

QC21.2
L578
v.1
1970

Theoretical Physics

An Advanced Text

Volume 1:

THEORY OF THE ELECTROMAGNETIC FIELD
THEORY OF RELATIVITY



Benjamin G. Levich

Institute of Electrochemistry
Academy of Sciences of the USSR, Moscow

69. 3. 10



For your professional use.

b10421476
109986

QC 21.8
L578
v.1
1970

Theoretical Physics

An Advanced Text

Volume 1:

THEORY OF THE ELECTROMAGNETIC FIELD
THEORY OF RELATIVITY



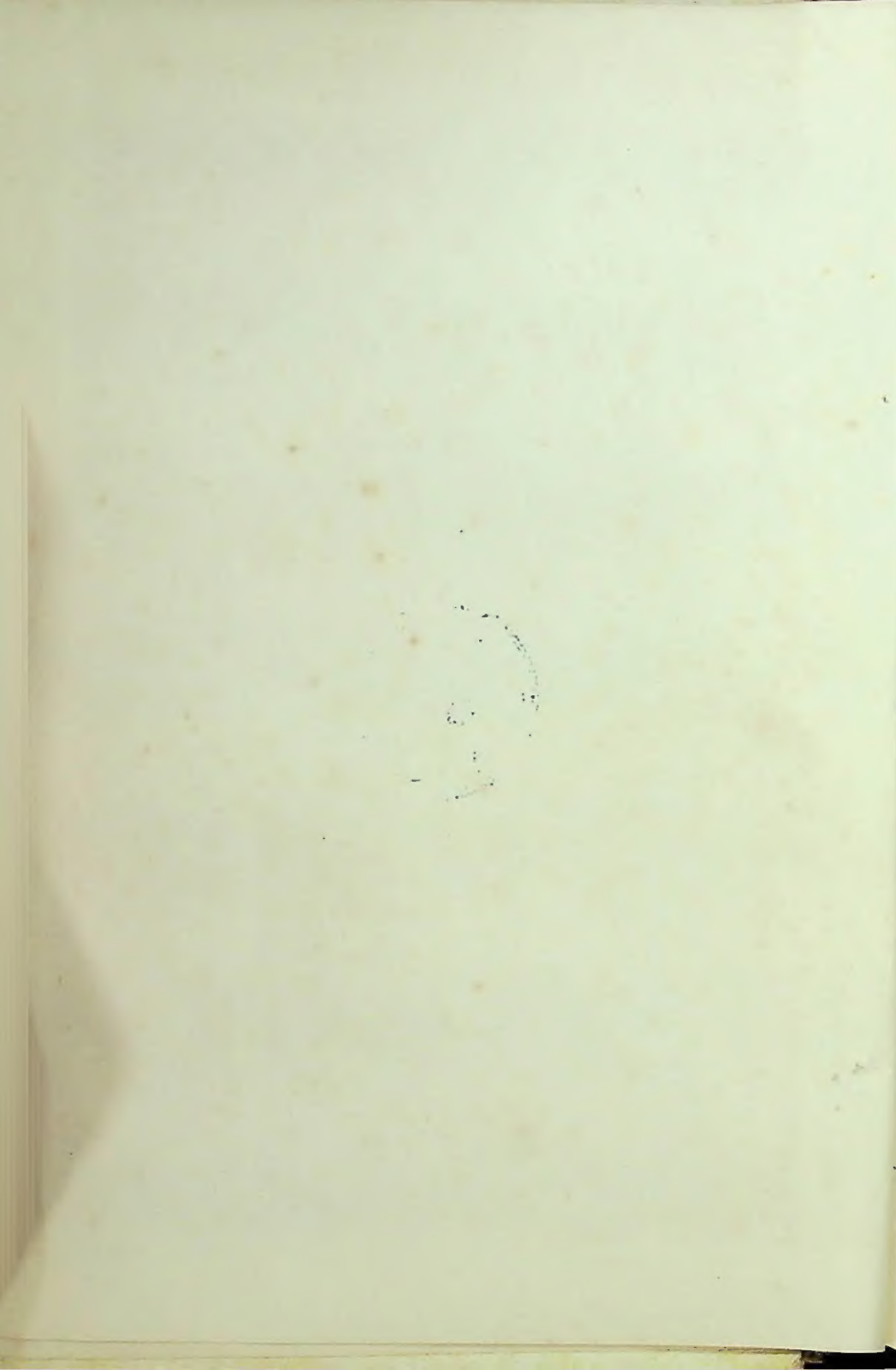
Benjamin G. Levich

Institute of Electrochemistry
Academy of Sciences of the USSR, Moscow

69. 3. 10



.b10421476
109986



QC21.
L578
v.1
1970

Theoretical Physics

An Advanced Text

Volume 1:

THEORY OF THE ELECTROMAGNETIC FIELD
THEORY OF RELATIVITY



Benjamin G. Levich

Institute of Electrochemistry
Academy of Sciences of the USSR, Moscow

69. 3. 10



For your professional use.

.b10421476
109986

© NORTH-HOLLAND PUBLISHING COMPANY, AMSTERDAM, 1970

All rights reserved: No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior permission of the Copyright owner.

Library of Congress Catalog Card Number: 68 54501

ISBN complete set: 0 7204 0176 3

Vol. 1: 0 7204 0177 1

Printed in The Netherlands

Title of the Russian edition:

KURS TEORETICHESKOJ FIZIKI

Russian edition published by:

IZDATELSTVO 'NAUKA', GLAVNAJA REDAKCIJA,

FIZIKO-MATEMATIČESKOJ LITERATURY (MOSKVA, 1969)

FOREWORD

The first Russian edition of 'Theoretical Physics', which appeared in 1962, has been widely used as a textbook.

Numerous comments from colleagues, lecturers and students have been taken into account in preparing this new edition, which is the first one in English and which will also appear as the second Russian edition.

The material has now been divided into 4 volumes covering the following subjects

Volume 1

- Part I Theory of the Electromagnetic Field
- Part II Theory of Relativity

Volume 2

- Part III Statistical Physics
- Part IV Electromagnetic Processes in Matter

Volume 3

- Part V Quantum Mechanics

Volume 4

- Part VI Quantum Statistics and Physical Kinetics

The rapid development of physics and the present wide interest in non-equilibrium and non-stationary processes has compelled us to expand the section on physical kinetics. It has also been transferred to the end of Volume 4 as it is practically impossible to expound this topic without using quantum mechanics.

Part IV — 'Electromagnetic Processes in Matter' — has been substantially revised. Interest in this field has increased recently, mainly in connection with the study of plasmas and plasma-like media, which now have sections devoted to them.

The methods of calculating electrostatic and direct-current fields, and other problems of classical electrodynamics in a medium, are covered very briefly as we have assumed that students will be able to consult the many monographs and handbooks on general physics, electrical- and radio-technology, and the equations of mathematical physics.

As for other modifications and additions, we should draw attention to the introduction of tensor notation, to new ideas in the theories of relativity and electromagnetic fields, the broadening of the introduction to the theory of probability, a brief presentation of the method of correlation functions in statistical physics, the exposition of the thermodynamic theory of ferromagnetism and the theory of propagation of electromagnetic waves in plasma. A number of paragraphs have been rewritten. We have tried to bring the content of the book even closer to the interests of present-day theoretical physics.

The general level of the book has been preserved and it is still intended to form an introduction to theoretical physics. Problems requiring the use of cumbersome or special mathematical apparatus are still excluded, and the most difficult sections are marked by an asterisk. These may be skipped at will, since there is no reference to them in the main text.

In conclusion we would like to express our gratitude to all those who helped us in preparing this book, in particular to A.M. Brodsky, A.M. Golovin, B.M. Grafov, R.R. Dogonadze, V.S. Krylov and especially V.S. Markin and V.V. Tolmachev. I.V. Savelyev discovered a number of misprints which have now been corrected.

L.D. Konkina helped us in editing the manuscript.

We are grateful to the readers and students who used the first Russian edition of the book for sending us their valuable comments which have been taken into account in this edition.

August 1970

FOREWORD TO THE FIRST RUSSIAN EDITION

The continuous development of theoretical physics and the regular expansion of its areas of application create increasing demand for textbooks and manuals.

The rapid development and the complexity of the most recent experimental methods of physical investigation, and the corresponding development and extension of the mathematical apparatus of theoretical physics, have meant that one man usually cannot combine the two methods of investigation. The end of the 19th century and particularly the 20th century therefore saw physicists divided into 'experimentalists' and 'theoreticians', the latter studying physical laws by means of the mathematical methods of theoretical physics.

Obviously, a background in theoretical physics is essential in the education of experimental as well as theoretical physicists.

The experimental and theoretical methods of physical investigation have penetrated into a number of branches of science related to physics (physical chemistry, biophysics, geophysics, astrophysics, and so on) and into technology (metal physics and metallurgical science, thermophysics, electrical technology, radiotechnology, computation, the instrument-making industry etc.). Workers in these branches of science and technology also need a certain minimum knowledge of theoretical physics.

The compilation of a modern textbook on theoretical physics is inevitably associated with certain logical and methodological difficulties. It is impossible at present to divide theoretical physics into classical and quantum parts so that it is also impossible to divide it into separate chapters and sections. For example, the exposition of statistical physics without taking into account the quantum properties of atomic systems is impossible, for it would mean that the general theory remained without practical application. In the theory of electromagnetic processes in matter one has of necessity to make use of the ideas of statistical physics, and so on. It may be that the maximum consistency of composition would be obtained if the book were founded on

quantum mechanics but this is completely inadmissible in a book intended as an introductory treatise. Quantum mechanics requires a certain preparedness and the student must be convinced of the necessity of renouncing obvious classical representations. Compromise solutions, which have justified themselves during many years of teaching theoretical physics at the Moscow Engineering-Physical Institute and Moscow State University, are therefore inevitable.

The following general principles have been applied.

(1) The book is written as an introduction to theoretical physics so that aspects requiring the use of cumbersome or special mathematical apparatus have not been included.

(2) As it is to be used for a systematic study of the subject the course is a unique whole and all material necessary for understanding the later sections is contained in the earlier ones.

(3) It would not be feasible to elucidate experimental facts in addition to problems concerning purely theoretical physics. However, physics is a single science, and an attempt to expound the theoretical aspects without taking experiment into account would be quite wrong. The reader is assumed to have some basic experimental knowledge from university courses in general and atomic physics so that we have confined ourselves to references and, in a few instances, to a schematic description of basic experiments.

(4) The acquaintance assumed with general courses in general and atomic physics has allowed us to rely on a certain (very restricted) knowledge of quantum mechanics in our treatment of statistical physics.

(5) Classical mechanics usually forms a separate course so that this topic has been omitted although detailed reference has been made to handbooks of mechanics.

(6) The book similarly does not cover hydrodynamics, aerodynamics, the theory of heat transfer, or problems related to electrical- and radio-technology.

(7) Detailed reference is made to mathematical manuals. The mathematical apparatus utilized, except in the sections marked by an asterisk, is covered by the usual courses in analysis. In the case of quantum mechanics, however, the mathematical apparatus has been included, since it is of a specific character and is not taught in traditional mathematical courses.

(8) As the book is intended as a systematic course in theoretical physics no attempt has been made to achieve the same level of accessibility in all sections. It is a well-known fact that a student's comprehension and assimilation of difficult material increases as a course progresses, and that this is also true for the associated mathematical apparatus. Moreover, experi-

mental physicists will constantly encounter new problems in quantum mechanics which can only be handled using advanced methods of treatment. The section on quantum mechanics (Part V) therefore deals with some topics having a more advanced character than those in other sections. The analysis of applications of the kinetic equations is similarly treated rather extensively.

The uniqueness of the book's objectives has affected the content of individual sections, so that some topics in modern physics have been included at the expense of more traditional material.

Part I contains the foundations of the theory of the electromagnetic field in a vacuum, based on the system of Maxwell-Lorentz equations. A basic knowledge of electromagnetism is assumed. The focus of attention is the theory of radiation and the motion of charged particles in external fields.

In Part II, devoted to the theory of relativity, a four-dimensional form of representation is adopted which not only corresponds to the spirit of the theory but also predominates in contemporary literature. The problems of dynamics in the theory of relativity are treated in some detail. A number of the most recent applications of the theory of relativity, particularly those related to nuclear physics, are covered here for the first time in a textbook.

Part III is a revised version of Levich's 'Introduction to Statistical Physics' and treats statistical physics and the fundamentals of statistical thermodynamics. Classical thermodynamics would require too much space, and did not seem indispensable.

Part IV contains the theory of electromagnetic processes in matter. Relatively little attention is paid to problems in theoretical electrical- and radio-technology. The phenomenological theory of electric and magnetic properties of matter is analyzed in some detail, and the notion of the physics of the plasma state of matter is given.

In Part V the basic ideas of present-day relativistic quantum mechanics are included as well as the traditional problems of non-relativistic quantum mechanics. Applications to solid-state theory are considered at length.

Part VI contains the essential concepts of physical kinetics, which are not usually presented in a general course on theoretical physics.

The experience of teaching theoretical physics shows that the greatest difficulties are often encountered not in understanding new physical ideas but in the actual mathematical treatments. All mathematical operations have therefore been performed in sufficient detail.

For convenience we have presented a brief derivation of those formulae of

vector analysis which are encountered throughout, as well as the necessary data on Fourier integrals and δ -function theory.

The numbering of formulae and sections starts afresh in each Part and references to appendices have been given Roman numerals.

The author hopes that the readers, after making themselves familiar with the foundations of theoretical physics expounded in this book, will be able to proceed to a more profound study using the many-volume treatise of Landau and Lifshitz. The scientific and educational ideas of their work were of great influence on the author, who is a disciple of Landau.

Parts I–IV and Part VI were written by B.G. Levich. Part V was written by Y.A. Vdovin and V.A. Myamlin under the general scientific guidance of B.G. Levich. Chapter XV * of Part V was written by A.I. Naumov.

The author expresses his gratitude to the colleagues who read the book and the manuscripts, and made a number of valuable remarks: B.M. Grafov, R.R. Dogonadze, V.A. Kiryanov, V.S. Krylov, V.S. Markin, V.P. Smilga, Y.A. Chizmadzhev and Y.I. Yalamov.

The creation of a textbook on theoretical physics sufficiently comprehensive in content and clear in presentation is a very complex task. The author is therefore conscious of the fact that shortcomings and errors will be discovered and would be grateful to receive an account of them which can be taken into consideration in the next edition of the book.

1962

* Chapter 13 of the English edition.

Theoretical Physics: Outline of Vols. 1—4

Volume 1 (for details see p. xv)

Part I Theory of the Electromagnetic Field

- Chapter 1 General theory of the electromagnetic field
- 2 The electrostatic field
 - 3 The quasistationary magnetic field
 - 4 The electromagnetic field of arbitrarily moving charges
 - 5 Radiation theory
 - 6 Electromagnetic field in a vacuum and electromagnetic wave scattering
 - 7 The motion of particles in electromagnetic fields

Part II Theory of Relativity

- Chapter 1 General principles of the theory of relativity
- 2 Relativistic mechanics
 - 3 Relativistic electrodynamics

Appendix I, II and III

Subject index

Volume 2

Part III Statistical Physics

- Chapter 1 The basic concepts of the theory of probability
- 2 The kinetic theory of gases

- 3 Statistical distribution
- 4 Statistical and phenomenological thermodynamics
- 5 Ideal gases
- 6 Systems of interacting particles
- 7 Crystals
- 8 The theory of fluctuations
- 9 Systems with a variable number of particles
- 10 Statistical distributions in quantum statistics and some of their applications

Part IV Electromagnetic Processes in Matter

- Chapter 1 Electromagnetic fields in matter
- 2 Electrostatics
 - 3 Direct electric current and the magnetic properties of matter
 - 4 Quasistationary electromagnetic fields
 - 5 High-frequency fields
 - 6 Matter in the plasma state

Appendix IV

Subject index

Volume 3

Part V Quantum Mechanics

- Chapter 1 The basic concepts of quantum mechanics
- 2 The Schrödinger equation
 - 3 The mathematical apparatus of quantum mechanics
 - 4 Motion in a centrally symmetric field
 - 5 The quasi-classical approximation
 - 6 The matrix form of quantum mechanics
 - 7 Perturbation theory
 - 8 Spin and identity of particles
 - 9 Applications of quantum mechanics to the consideration of the properties of atomic and nuclear systems
 - 10 The theory of diatomic molecules
 - 11 Scattering theory

- 12 The method of second-quantization and radiation theory
- 13 Relativistic quantum mechanics
- 14 Some problems of quantum electrodynamics
- 15 Fundamentals of the theory of elementary particles

Subject index

Volume 4

Part VI Quantum Statistics and Physical Kinetics

- Chapter 1 Quantum statistics
- 2 Physical kinetics
- 3 Kinetic theory of gases and gas-like systems
- 4 Time correlative function method and Onsager's theory
- 5 Solid-state theory

Subject index



Contents of Volume 1

Part I Theory of the electromagnetic field

Chapter 1 General theory of the electromagnetic field

§ 1	Problems of theoretical physics	1
2	The determination of the vector field from its integral characteristics	3
3	Charges and particles	9
4	The field of charges at rest	11
5	The equation of continuity	16
6	The electromagnetic field of charges moving with a constant velocity	18
7	The electromagnetic field of moving charges. The general case	23
8	Maxwell—Lorentz system of equations	28
9	The displacement current	30
10	The electromagnetic field potentials	34
11	Gauge invariance of the potentials	37
12	Energy conservation law of the electromagnetic field	40
13	Momentum conservation law of the electromagnetic field	44

Chapter 2 The electrostatic field

14	The electrostatic field	49
15	The electrostatic field of a system of point charges	52
16	Quadrupole moment	57
17	Work and energy in an external electrostatic field	62
18	The interaction energy of a system of charges and the electrostatic field energy	65

Chapter 3 The quasistationary magnetic field

§19 The field of a system of charges undergoing a slow quasistationary motion	69
20 The field of a point charge undergoing a slow uniform motion	76
21 The field of a system of charges undergoing quasistationary motion at a large distance from the system	79
22 The magnetic moment	82

Chapter 4 The electromagnetic field of arbitrarily moving charges

23 The electromagnetic field of a system of arbitrarily moving charges	85
24 General solution of D'Alembert's equation in the form of retarded potentials	94
25 The field of a point charge moving arbitrarily	102

Chapter 5 Radiation theory

26 The potentials of the electromagnetic field at a large distance from the emitter in the dipole approximation	106
27 The electromagnetic field of dipole radiation at a large distance from the emitter	112
28 Dipole radiation of simple systems	116
29 Radiation reaction	121
30 Line width of emitted radiation	125
31 Quadrupole and magnetic dipole radiation	129
32 General case of electromagnetic radiation. The spectral decomposition of fields. The radiation zone and induction zone. Effect of the proper retardation	134

Chapter 6 Electromagnetic field in a vacuum and electromagnetic wave scattering

33 The propagation of electromagnetic waves at a large distance from the emitter	142
34 Plane wave polarization	149
35 Interference and the formation of wave packets	152
36 Scattering of the electromagnetic waves by a free charge and by a bound electron	157
37 Absorption of radiation	162
38 Canonical form of the field equations	164

Chapter 7 The motion of particles in electromagnetic fields

§ 39 The motion of charged particles in constant electric and magnetic fields	174
40 The motion of charged particles in slowly varying magnetic fields	182
41 The Lagrangian and Hamiltonian for a particle moving in an electromagnetic field	187
42 The motion of a system of two charged particles and the radiation from them	189
43 The scattering of particles and associated bremsstrahlung	196

Part II Theory of relativity*Chapter 1 General principles of the theory of relativity*

§ 1 The creation and significance of the theory of relativity	215
2 Galilean transformations	216
3 Attempts to determine an absolute velocity	219
4 Postulates of the Einstein theory of relativity	222
5 The Lorentz transformation	224
6 Consequences of the Lorentz transformation. Space and time intervals	228
7 Einstein's law of addition of velocities and angular transformations	234
8 Simultaneity, short-range action and action at a distance	237
9 Absolute values in the theory of relativity. Intervals and proper time	238
10 The invariance of physical laws under Lorentz transformations. Four-dimensional formulation of the theory of relativity	242
11 Four-dimensional vectors and tensors. Four-dimensional velocity and acceleration	248

Chapter 2 Relativistic mechanics

12 The dynamical equations of a material point	256
13 Momentum, energy and mass in relativistic mechanics	260
14 Lagrange's equations; the Lagrangian and Hamiltonian	267

§ 15	The mechanics of a system of particles in the theory of relativity	268
16	The energy-momentum conservation law in nuclear physics	274
17	The theory of collisions between relativistic particles. Compton effect	286
 <i>Chapter 3 Relativistic electrodynamics</i>		
18	Charge conservation, the four-dimensional current and the equation of continuity	292
19	The relativistically invariant formulation of the equations of the electromagnetic field potentials	294
20	The field of a moving charge	296
21	The electromagnetic field tensor and Maxwell's equations	304
22	Some applications: Doppler effect, Mössbauer effect, observation of rapidly moving bodies, transformation of angles, intensities and cross sections	307
23	The Lorentz force; the Lagrangian and the Hamiltonian for a particle moving in an electromagnetic field	325
24	The motion of particles in constant electric and magnetic fields	331
25	A system of weakly interacting charged particles	340
26	The radiation emitted by a moving charge	350
Appendix I Vector analysis		356
Appendix II The Fourier integral		375
Appendix III The delta-function and its properties		379
Subject index		387

PART I

THEORY OF THE
ELECTROMAGNETIC FIELD



General Theory of the Electromagnetic Field

§1. Problems of theoretical physics

Physics is first of all an experimental science. However, in the work of Newton and other founders of contemporary physics, mathematical methods were successfully applied to obtain the quantitative formulation of physical laws, at first mainly in mechanics.

In the last century the application of mathematical methods to physics was so extensive that there arose the particular branch of physics known as theoretical physics. Problems facing theoretical physics are of two kinds:

1. The expression of physical laws in the form of quantitative relations, and the establishment of underlying correlations among experimental facts.
2. The application of mathematical methods of investigation to find new physical laws and the prediction of new, as yet not experimentally observed connections between physical phenomena.

Thus, theoretical physics is, in its methods, a mathematical science, and, in its content, a physical science.

From the above it is clear that it is in theoretical physics that general theoretical views concerning the essence of various physical processes are embodied and made complete.

It is best to illustrate this by a simple example. Investigators who established experimentally the planetary atomic model and the presence of dis-

crete allowed energy values in atoms and similar facts, gave a paramount contribution to theoretical physics. However, theoretical physics could not confine itself to the qualitative model representations of the atomic structure. From this it follows that theoretical physics seeks to formulate the most general quantitative physical laws expressing the essence of as wide a range of phenomena as possible. Mechanical laws (Newtonian laws), the laws of the electromagnetic field (Maxwell-Lorentz equations), and the laws of quantum mechanics etc. serve as examples.

Logical reasoning cannot be the basis of general physical laws. Only experimental facts can be this. Hence the most general quantitative relations of theoretical physics are not 'derived', but represent a generalized formulation of observed physical regularities. On the other hand, as we shall see in concrete examples, in a number of cases the quantitative expression of physical laws appeared as a result of scientific prediction.

Having at its disposal a quantitative formulation of general physical laws, theoretical physics can undertake the second part of its program — the establishment of new laws and relations by means of mathematical methods. In this way theoretical physics has achieved such great successes that in comparison with them even such examples of scientific prediction as Leverrier's discovery of the planet Neptune in the nineteenth century appear to be of minor importance.

As examples one can point to the discovery of displacement current by Maxwell and the consequent establishment of the electromagnetic nature of light; the foundation of the theory of relativity by Einstein and, in particular, establishment of the mass-energy relation; the prediction by quantum mechanics (founded by de Broglie, Schrödinger and Heisenberg) of the existence of the wave properties of microparticles — electrons, protons, etc.; the prediction by Dirac's theory of the existence and properties of the positron and other antiparticles, and so forth. The role of theoretical physics in the recent development of nuclear physics and in the application of atomic energy is well-known.

It is necessary to stress that the methods of calculation of theoretical physics are of a special character. Theoretical physics is not a branch of mathematics. In theoretical physics one does not try to find the exact physical laws defining the behaviour of even relatively simple systems. An exact calculation of all possible effects and interrelations would make even the most simple problems insoluble. The necessity of taking into account essential relations and disregarding unessential ones is borne in mind in every stage of investigation in theoretical physics. The relations and equations of theoretical physics are so complicated that one must practically always proceed by

approximate calculations. In order to find out which approximations are possible and advisable and which ones are unjustifiable and physically senseless, one must often proceed from available experimental data. At the same time, formulae and relations which in principle cannot be checked experimentally are not considered at all in theoretical physics. All efforts of theoretical as well as experimental physics are directed to explaining objectively existing relations, i.e. the physical laws of nature.

A physical theory explaining known facts but unable to predict new ones is always considered unsatisfactory. On the other hand, the highest appraisal of the validity of a physical theory is the experimental confirmation of the facts predicted by it. In its turn, the elucidation of new phenomena observed experimentally serves as a stimulus for further development of theoretical physics. Thus, experimental and theoretical physics make a single and inseparable whole.

§ 2. The determination of the vector field from its integral characteristics

We shall see in what follows that the state of an electromagnetic field is specified by its vector characteristics — i.e. the strengths of the electric and magnetic fields. For this reason, in putting forward the general theory and in solving definite problems in the theory of the electromagnetic field, extensive use is made of a specific mathematical apparatus, the so-called vector analysis.

A description of the electromagnetic field which is not based on vector analysis is possible in principle, but would require very cumbersome calculations and complicated transformations. Hence the subsequent exposition is carried out solely on the basis of vector analysis. Although we assume that the reader is familiar with its fundamentals, Appendix I gives a brief derivation of all the formulae and transformations which are to be encountered.

We shall analyse here an important problem of the mathematical theory of the arbitrary vector field. The importance of this problem in the theory of the electromagnetic field lies in the fact that the general scheme of calculation of the field theory is constructed according to the calculation of the arbitrary vector field presented below.

Let there be a vector field $\mathbf{a}(\mathbf{r})$ over all space. Some assumptions, which will be mentioned below, will be made about the behaviour of the vector $\mathbf{a}(\mathbf{r})$ at infinitely distant points of space.

Assume that at every point of space the integral characteristics of the field, the vector flux $\oint \mathbf{a} \cdot d\mathbf{S}$ and the vector circulation $\oint \mathbf{a} \cdot d\mathbf{l}$, are given. We shall see in what follows that for electromagnetic fields it is just these character-

istics which contain the quantities which can be directly measured experimentally. We shall show that, if these field characteristics are given, then the vector field $\mathbf{a}(\mathbf{r})$ can itself be found. If $\oint \mathbf{a} \cdot d\mathbf{S} = \int f(\mathbf{r}) dV$, where $f(\mathbf{r})$ is a known function of coordinates, then on the basis of the Gauss-Ostrogradsky theorem

$$\oint \mathbf{a} \cdot d\mathbf{S} = \int \nabla \cdot \mathbf{a} dV = \int f(\mathbf{r}) dV, \quad (2.1)$$

and in view of the arbitrary character of the integration range over a volume V , we have

$$\nabla \cdot \mathbf{a} = f(\mathbf{r}). \quad (2.2)$$

Thus, the definition of the vector flux through a closed surface at every point of space is equivalent to the definition of the divergence of the vector.

Further, on the basis of Stokes' theorem,

$$\oint \mathbf{a} \cdot d\mathbf{l} = \oint (\nabla \times \mathbf{a}) \cdot d\mathbf{S} = \int \boldsymbol{\omega}(\mathbf{r}) \cdot d\mathbf{S}, \quad (2.3)$$

where $\boldsymbol{\omega}(\mathbf{r})$ is a known vector function of coordinates. Hence

$$\nabla \times \mathbf{a} = \boldsymbol{\omega}(\mathbf{r}). \quad (2.4)$$

The definition of the vector circulation is equivalent to the definition of its curl.

We shall show how the vector field $\mathbf{a}(\mathbf{r})$ can be found if the divergence and curl of the vector \mathbf{a} are known over all space.

We resolve the field \mathbf{a} into two fields: $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2$, so that the following relations hold:

$$\nabla \cdot \mathbf{a}_1 = f(\mathbf{r}), \quad (2.5)$$

$$\nabla \times \mathbf{a}_1 = 0, \quad (2.5')$$

$$\nabla \cdot \mathbf{a}_2 = 0, \quad (2.6)$$

$$\nabla \times \mathbf{a}_2 = \boldsymbol{\omega}(\mathbf{r}). \quad (2.6')$$

The vector field \mathbf{a}_1 is vortex-free; its lines begin and end in the sources and sinks whose intensity is given by the function $f(\mathbf{r})$. The vector field \mathbf{a}_2 has no sources and no sinks, and is a solenoidal field.

Let us begin with the consideration of the field \mathbf{a}_1 . Since the field \mathbf{a}_1 is

vortex-free, the vector \mathbf{a}_1 can be written in the form of the gradient of a certain auxiliary scalar function

$$\mathbf{a}_1 = \nabla \varphi(\mathbf{r}), \quad (2.7)$$

where $\varphi(\mathbf{r})$ is a function called the scalar potential. Substituting eq. (2.7) into (2.5), we find $\nabla \cdot \nabla \varphi = f(\mathbf{r})$, or

$$\nabla^2 \varphi = f(\mathbf{r}), \quad (2.8)$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

is the Laplacian.

Eq. (2.8) represents a second order partial differential equation, called Poisson's equation. We shall obtain its general solution in § 24. Here we shall give only the final result and convince ourselves of the fact that this solution satisfies eq. (2.8). It turns out that the solution of Poisson's equation has the form

$$\begin{aligned} \varphi(\mathbf{r}) &= -\frac{1}{4\pi} \int \frac{f(\mathbf{r}_0) dV_0}{|\mathbf{r} - \mathbf{r}_0|} = \\ &= -\frac{1}{4\pi} \int \frac{f(x_0, y_0, z_0) dx_0 dy_0 dz_0}{\sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}}, \end{aligned} \quad (2.9)$$

where x, y, z are the coordinates of the point of observation, i.e. of that point at which the value of the function φ is sought, while x_0, y_0, z_0 are the variables of integration. The quantity

$$|\mathbf{r} - \mathbf{r}_0| = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}$$

represents the distance from the point \mathbf{r}_0 to the point of observation \mathbf{r} . Substituting (2.9) into (2.8) one can easily verify that the expression given indeed satisfies the initial equation:

$$\begin{aligned}
\nabla^2 \varphi &= - \nabla^2 \frac{1}{4\pi} \int \frac{f(x_0, y_0, z_0) dx_0 dy_0 dz_0}{|\mathbf{r} - \mathbf{r}_0|} = \\
&= - \frac{1}{4\pi} \int f(x_0, y_0, z_0) dV_0 \nabla^2 \frac{1}{|\mathbf{r} - \mathbf{r}_0|} = \\
&= \int f(x_0, y_0, z_0) \delta(x - x_0) \delta(y - y_0) \delta(z - z_0) dx_0 dy_0 dz_0 = \\
&= f(x, y, z).
\end{aligned}$$

Here we have made use of the properties of the δ -function and the function $1/|\mathbf{r} - \mathbf{r}_0|$ (see Appendix III). The operator ∇^2 denotes differentiation with respect to the coordinates \mathbf{r} and can be brought under the integration sign with respect to the coordinates \mathbf{r}_0 .

If the integration is carried out over all space, it is necessary for the existence and convergence of the integral (2.9) that the integrand $f(\mathbf{r}_0)$ should satisfy the obvious requirement

$$|f(\mathbf{r}_0) \cdot r_0^{2+\lambda}| < A \quad \text{as} \quad r_0 \rightarrow \infty, \quad (2.10)$$

where A is a finite quantity and $\lambda > 0$. In other words, the function $f(\mathbf{r}_0)$ must decrease more rapidly than the function $1/r_0^2$ as $r_0 \rightarrow \infty$. When this condition is fulfilled the integral (2.9) converges, and the function $\varphi(r)$ decreases as its argument increases indefinitely according to the law

$$|\varphi(r)| < 1/r. \quad (2.11)$$

Assuming the condition (2.10) to be fulfilled, we can state that (2.9) represents the solution of eq. (2.8).

Knowing the function $\varphi(\mathbf{r})$ and making use of the definition (2.7), we find

$$\mathbf{a}_1(\mathbf{r}) = \nabla \varphi(\mathbf{r}) = - \frac{1}{4\pi} \nabla \int \frac{f(\mathbf{r}_0) dV_0}{|\mathbf{r} - \mathbf{r}_0|}. \quad (2.12)$$

We now pass on to the determination of the vector $\mathbf{a}_2(\mathbf{r})$. The vector \mathbf{a}_2 is of a solenoidal character and, consequently, can be written in the form of the curl of a certain auxiliary vector $\mathbf{A}(\mathbf{r})$:

$$\mathbf{a}_2 = \nabla \times \mathbf{A}(\mathbf{r}) \quad (2.13)$$

The vector function $\mathbf{A}(\mathbf{r})$ is called the vector potential. From the definition (2.13) it is clear that eq. (2.6) is satisfied automatically:

$$\nabla \cdot \nabla \times \mathbf{A}(\mathbf{r}) = 0.$$

To determine the vector potential \mathbf{A} completely it is necessary in addition to give the value of its divergence $\nabla \cdot \mathbf{A}$, which for the present remains arbitrary. We assume

$$\nabla \cdot \mathbf{A} = 0. \quad (2.14)$$

Somewhat later we shall verify the fact that the above assumption does not restrict the general character of our reasoning. Substituting (2.13) into (2.6'), we have

$$\nabla \times [\nabla \times \mathbf{A}(\mathbf{r})] = \nabla [\nabla \cdot \mathbf{A}(\mathbf{r})] - \nabla^2 \mathbf{A}(\mathbf{r}) = \boldsymbol{\omega}(\mathbf{r}).$$

Taking into account (2.14), we find

$$\nabla^2 \mathbf{A}(\mathbf{r}) = -\boldsymbol{\omega}(\mathbf{r}) \quad (2.15)$$

or, in the scalar form,

$$\nabla^2 A_x = -\omega_x(\mathbf{r}), \quad \nabla^2 A_y = -\omega_y(\mathbf{r}), \quad \nabla^2 A_z = -\omega_z(\mathbf{r}).$$

The components of the vector potential satisfy the same equations as the scalar potential φ . Their solutions read:

$$A_x = \frac{1}{4\pi} \int \frac{\omega_x(\mathbf{r}_0)}{|\mathbf{r} - \mathbf{r}_0|} dV_0, \quad (2.16)$$

$$A_y = \frac{1}{4\pi} \int \frac{\omega_y(\mathbf{r}_0)}{|\mathbf{r} - \mathbf{r}_0|} dV_0, \quad (2.16')$$

$$A_z = \frac{1}{4\pi} \int \frac{\omega_z(\mathbf{r}_0)}{|\mathbf{r} - \mathbf{r}_0|} dV_0 \quad (2.16'')$$

If the functions ω_x , ω_y and ω_z satisfy the same conditions at infinity as those which must be satisfied by the function $f(\mathbf{r}_0)$ in (2.9), then the integrals in the expressions (2.16)–(2.16'') converge. In this case formulae (2.16)–

(2.16'') determine the vector potential \mathbf{A} . Knowing the vector potential \mathbf{A} , one can find the vector \mathbf{a}_2 by a simple differentiation:

$$\mathbf{a}_2(\mathbf{r}) = \nabla \times \left[\frac{1}{4\pi} \int \frac{\boldsymbol{\omega}(\mathbf{r}_0)}{|\mathbf{r} - \mathbf{r}_0|} dV_0 \right]. \quad (2.17)$$

By differentiation one can also show that the vector potential obtained satisfies the condition (2.14).

Thus, the vector field $\mathbf{a}(\mathbf{r})$ is completely determined if the values of its divergence, $f(\mathbf{r}_0)$ and curl, $\boldsymbol{\omega}(\mathbf{r}_0)$ in all space are given:

$$\begin{aligned} \mathbf{a}(\mathbf{r}) &= \mathbf{a}_1 + \mathbf{a}_2 = \nabla \varphi + \nabla \times \mathbf{A} = \\ &= -\frac{1}{4\pi} \nabla \left[\int \frac{f(\mathbf{r}_0)}{|\mathbf{r} - \mathbf{r}_0|} dV_0 \right] + \frac{1}{4\pi} \nabla \times \left[\int \frac{\boldsymbol{\omega}(\mathbf{r}_0)}{|\mathbf{r} - \mathbf{r}_0|} dV_0 \right]. \end{aligned} \quad (2.18)$$

Since the divergence and curl of the vector \mathbf{a} are uniquely related to the flux and circulation of this vector, it can also be stated that the vector field $\mathbf{a}(\mathbf{r})$ is completely determined by the flux and circulation of this vector.

We shall now dwell on the important problem of the possibility of choosing $\nabla \cdot \mathbf{A}$ in the form of (2.14). The vector potential \mathbf{A} is not given uniquely by the definition (2.13). One can add to it the gradient of an arbitrary function ψ , i.e. it can be assumed that

$$\mathbf{A}' = \mathbf{A} + \nabla \psi. \quad (2.19)$$

We have, obviously,

$$\nabla \times \mathbf{A}' = \nabla \times \mathbf{A}.$$

Thus, the addition of the gradient of an arbitrary function ψ to \mathbf{A} leads to the previous value of the vector \mathbf{a} .

Let, contrary to the condition (2.14), $\nabla \cdot \mathbf{A} \neq 0$. Then it is always possible to transform to a new vector potential \mathbf{A}' according to formula (2.19). For it we have

$$\nabla \cdot \mathbf{A}' = \nabla \cdot \mathbf{A} + \nabla \cdot \nabla \psi = \nabla \cdot \mathbf{A} + \nabla^2 \psi.$$

It is always possible, without restricting the general character of the reasoning, to choose the arbitrary function ψ in such a way that for any $\nabla \cdot \mathbf{A} \neq 0$ the equality

$$\nabla^2 \psi = -\nabla \cdot \mathbf{A}$$

will hold. This means that it can always be assumed that

$$\nabla \cdot \mathbf{A}' = 0$$

and, consequently, that condition (2.14) is of a general character.

Finally, it is easy to show that the expression obtained for \mathbf{a} is the only solution of eqs. (2.2)–(2.4) *.

The expression which we have found for the vector field \mathbf{a} , depending on the values of its divergence, $f(\mathbf{r})$, and curl, $\omega(\mathbf{r})$, is not connected with any assumptions about the physical meaning and character of the quantities under consideration. At the same time, it represents the prototype of those calculations which one usually has to make in the theory of the electromagnetic field in order to find the electric and magnetic fields.

§3. Charges and particles

According to contemporary ideas, there are in nature elementary particles and systems which have a complex structure made up of elementary particles, e.g. atoms and molecules. Elementary particles and systems consisting of a relatively small number of elementary particles — individual atoms and molecules — are called microparticles and microsystems.

At present a large number of elementary particles, exceeding 20, is known. The interrelations between elementary particles are very far from the simple scheme which was adopted in physics a relatively short time ago, when only two elementary particles, the proton and the electron, were known.

We shall acquaint ourselves with the basic properties of microparticles and microsystems in what follows, and mainly in Part V of the book. The most profound problems concerning the structure and properties of elementary particles are not as yet elucidated in contemporary physics, and a number of established principles are so complex that we cannot present them within the framework of this book.

* See, for example, N. E. Kochin, *Vektornoe ischislenie i nachala tenzornovo ischisleniya* (Vector calculus and the principles of tensor calculus) (Izdatelstvo AN SSSR, 1965) p. 213; H. Lass, *Vector and tensor analysis* (McGraw-Hill, New York, 1950) p. 119.

In Part I of the book some properties of microparticles and microsystems will be considered in the approximation of classical physics. What this consists of and the limits of its applicability will be seen from what follows.

The most important characteristic of all microparticles is the law of interaction between them. Microparticles can interact when they are at a distance from each other. At present it is known that there are several different kinds of interaction between microparticles: the electromagnetic, gravitational and nuclear interactions. Each kind of interaction is associated with a definite characteristic of the particle. In Part I of this book we shall be interested solely in the electromagnetic interaction. This is understood to be a well defined interaction force between certain particles, the character of which will be considered below. The electromagnetic interaction does not depend on the masses of the particles, which determine their gravitational interaction. It turns out that in the simplest case of identical particles at rest with respect to each other the force of interaction is determined only by the distance between them and the unique characteristic of the particles called their charge.

The law of interaction between particles at rest with respect to each other is expressed by the well-known formula

$$\mathbf{F} = \frac{e_1 e_2}{r^3} \mathbf{r},$$

where \mathbf{F} is the force, \mathbf{r} is the distance between the particles, and the e_i are their charges. This is the so-called Coulomb law.

The charge of a particle of given type is one of its fundamental characteristics. The force of interaction between particles of a given type — electrons, protons and so on — is always a force of repulsion between them. The interaction between particles of different types can have the character of repulsion as well as attraction. The convention is adopted that electrons are negatively charged, and that protons are positively charged. The sign of the charge of other charged elementary particles — muons, pions, kaons and hyperons — is determined with respect to electrons and protons in the following way: particles having a charge of the same sign repel each other, whereas oppositely charged particles attract each other.

Neutrons, neutrinos and neutral π -mesons may serve as examples of neutral particles. No charge differing from zero can be ascribed to any neutral particle.

A striking feature of the charge is the fact that it has one and the same absolute value for all elementary particles. In the CGSE system, which we shall use in what follows, the elementary charge is equal to

$$|e| = 4.77 \times 10^{-10} \text{ g}^{1/2} \text{ cm}^{3/2} \text{ sec}^{-1}.$$

Another property of the charge, expressing its fundamental importance as a characteristic of particles, is that it is conserved. In all processes occurring in nature the algebraic sum of charges does not change (the charge conservation law). The charge conservation law is one of the most important laws of nature.

Most bodies in terrestrial conditions are made of atoms and molecules — quasi-neutral systems for which the positive charge of the nucleus is equal to the negative charge of the electron shell. When an atom is ionized, i.e. when it becomes charged, it loses one or more electrons. Because of the charge conservation law, on ionization there arises a positive ion with a charge $N|e|$ and N electrons each with a charge $-|e|$, where N is an integer. When an extra electron is added to an atom, the latter may be transformed into a negative ion with the charge $-|e|$. Thus the charge of any system is an integer multiple of the elementary charge e .

In the microscopic classical field theory we shall study the behaviour of systems consisting of a relatively small number of particles, for example, individual electrons or protons, ions and so on. We shall assume individual elementary particles to have no extension and to move according to the laws of classical mechanics. We shall not be interested in the internal structure of elementary particles. As will be clear from what follows, such an idealization appears to be too crude in a number of cases. The laws of classical physics have a restricted applicability to microsystems, and sometimes are not applicable to them at all. In particular, they are unsuitable for the consideration of phenomena taking place in a very small region of space near the charge. Hence in a number of cases, which will be considered later, our simplifying assumptions will lead to difficulties and contradictions.

In quantum mechanics (Part V) the concepts of the laws determining the motion and properties of microscopic particles will essentially be developed and improved.

Since for the present we shall be interested only in the properties of particles associated with their electromagnetic interaction, we shall simply speak about the interaction of charges.

§ 4. The field of charges at rest

Let there be fixed charges e_i at certain points of space \mathbf{r}_i . In the region of space near this set of charges we put a charge ϵ so small that the change in the

properties of the system caused by ϵ can be disregarded. We shall call such a charge the test charge. Observing the test charge, we shall find out that at each point of space \mathbf{r} it is acted upon by a force \mathbf{F} proportional to the value of the charge ϵ , i.e.

$$\mathbf{F} = \epsilon \mathbf{E}(\mathbf{r}). \quad (4.1)$$

Strictly speaking, a force \mathbf{F} acts on the test charge located at any distance from the fixed charges. However, since the value of the force decreases (see below) rapidly with increasing distance, the action of the force \mathbf{F} is manifested practically only in the vicinity of the charges. That region of space in which a force \mathbf{F} acts on the test charge will be called by us the region of the electric field of charges at rest or the electrostatic field. Sufficiently distant regions of space in which the force \mathbf{F} becomes negligibly small will be assumed, approximately, to be infinitely distant, and it will be considered that there is no field in them.

Since the test charge does not affect the properties of the field of the system of charges, a vector \mathbf{E} characterizes the properties of this field. We shall henceforth call \mathbf{E} the electric field strength or, briefly, the electric field.

Investigating the force \mathbf{F} acting on the test charge one can determine the value of the vector \mathbf{E} at every point of the field and thus establish the properties of the electric field of the fixed charges. It turns out that it is possible to find certain general properties of the fields of fixed charges which do not depend on the exact nature of the disposition or the values of the charges.

Experiment shows that the electric field of a system of charges at rest possesses additive properties: the strength of the total electric field produced by several charges is equal to the vector sum of the fields \mathbf{E}_i produced by each charge, i.e.

$$\mathbf{E} = \sum \mathbf{E}_i. \quad (4.2)$$

This most important property of electric fields is usually called the property of superposition.

Investigating the motion of the test charge one can find the vector lines of the electrostatic field \mathbf{E} .

Knowing the distribution of the field \mathbf{E} in space, one can determine the distribution and interaction of the charges producing it. The introduction of a field as a mathematical method may seem a natural convenient way of describing an interaction. We shall see below that in fact this is not so and that the field is as real as the particles. One can ascribe to the field the same

characteristics as to particles — energy, momentum, mass and so on. Moreover, we shall show that spatially separated particles cannot act directly on each other (the so-called long-range action). The particle changes the state of field in the immediate vicinity of itself. This change in the state of field, a perturbation, is propagated from point to point and reaches the other particle. Such is the concept of the field action, or the theory of short-range action. The theory of short-range action will be considered in more detail in §24 of Part I and §8 of Part II.

It turns out that the work performed by the electrostatic field on the test charge as the latter is displaced from a point \mathbf{r}_1 to a point \mathbf{r}_2 does not depend on the path over which this displacement takes place. This means that the work of displacement of the charge over a closed path is equal to zero, i.e.

$$W = e \oint \mathbf{E} \cdot d\mathbf{l} = 0.$$

Thus, for the lines of the electrostatic field the equality

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0 \quad (4.3)$$

always holds in integration over an arbitrary closed contour.

In connection with what was said earlier we pass over from the equality (4.3) to the differential characteristic of the field. For this we make use of Stokes' theorem, which gives

$$\oint \mathbf{E} \cdot d\mathbf{l} = \int (\nabla \times \mathbf{E}) \cdot d\mathbf{S} = 0. \quad (4.4)$$

Since the surface of integration in (4.4) is arbitrary, it follows from (4.4) that

$$\nabla \times \mathbf{E} = 0. \quad (4.5)$$

This formula shows that the electrostatic field is vortex-free. In it there are no closed field lines. Consequently, there must exist sources and sinks in which the lines of the field begin and end.

Experiment shows that in the electrostatic field the sources and sinks of the field lines are electric charges. It is assumed that the lines of the field originate on positive charges and end on negative ones. Since charges are the sources and sinks, the flux of the vector \mathbf{E} across any closed surface surrounding each charge is different from zero. If a certain amount of charge $\sum e_i$, where the summation denotes algebraic summation over all the charges, lies within a surface of integration S , then the flux of the vector \mathbf{E} must be pro-

portional to this sum, i.e.

$$\oint \mathbf{E} \cdot d\mathbf{S} = \text{const} \sum e_i. \quad (4.6)$$

This statement is called Gauss' theorem.

We stress that within the framework of our approximation the charges have a point character, and the value of the charge of any system is a multiple of the elementary charge.

To simplify the mathematical operations we take an important step and to pass over from the discrete, discontinuous distribution of charges to the continuous one.

Let there be in a certain small volume δV a sufficiently large number of charges. Since the charges are located at small distances from each other it is convenient for the mathematical description of them to replace the true distribution of the discrete point charges by a fictitious continuous distribution. Namely, replacing the volume δV by an infinitely small volume dV and assuming that an infinitely small charge de is confined in the infinitely small volume, one can write that

$$de = \rho dV,$$

where $\rho = de/dV$ is the charge density, i.e. the ratio of the charge to the volume it occupies at the given point of space. For charges at rest, ρ is a continuous function of the position $\rho(\mathbf{r})$.

It should be emphasized that the transition to the continuous distribution is of a purely mathematical character. It should not be confused with an analogous operation with which we shall become acquainted in Part IV where the electromagnetic processes in matter will be studied.

The connection between the mathematical description of the discrete distribution of point charges and the continuous function $\rho(\mathbf{r})$ can be established by means of the δ -function (see Appendix III). Namely, since the total charge in an arbitrary volume can be expressed in the form

$$e = \int \rho(\mathbf{r}) dV = \sum e_i,$$

where the summation is carried out over all the charges which are in the volume V , we can write

$$\rho(\mathbf{r}) = \sum_i e_i \delta(\mathbf{r} - \mathbf{r}_i),$$



where \mathbf{r}_i is the radius vector of the i th charge. Indeed, substituting this value $\rho(\mathbf{r})$, we have

$$\int \rho(\mathbf{r}) dV = \sum_i e_i \int \delta(\mathbf{r} - \mathbf{r}_i) dV = \sum_i e_i.$$

In particular, the charge density corresponding to one charge located at the point \mathbf{r}_0 can be written in the form

$$\rho(\mathbf{r}) = e \delta(\mathbf{r} - \mathbf{r}_0).$$

The importance of the introduction of the continuous charge density lies in the fact that by its introduction the field itself as well as the charge distribution are described by continuous functions of position.

Making use of the definition of charge density, we can write Gauss' theorem in the form

$$\oint \mathbf{E} \cdot d\mathbf{S} = \text{const} \int \rho dV, \quad (4.7)$$

where the integral on the right is taken over the volume confined by the surface S . According to the Gauss-Ostrogradsky theorem,

$$\oint \mathbf{E} \cdot d\mathbf{S} = \int \nabla \cdot \mathbf{E} dV. \quad (4.8)$$

Hence from formulae (4.7) and (4.8) we obtain

$$\nabla \cdot \mathbf{E} = \text{const} \rho. \quad (4.9)$$

Formula (4.9) determines the field divergence at every point of space. The value of the factor of proportionality in formula (4.9) can be determined only from experimental data (for example, from the Coulomb law).

In the CGSE system, which we shall use, this constant is equal to 4π , so that

$$\nabla \cdot \mathbf{E} = 4\pi\rho. \quad (4.10)$$

For reasons which will be explained below, we shall call eqs. (4.5) and (4.10) Maxwell's equations for the electrostatic field or, briefly, the equations of electrostatics.

Since the electrostatic field is vortex-free, in correspondence with the gen-

eral methods of describing the vector field one can introduce a scalar φ called the electrostatic potential and determined by the relation

$$\mathbf{E} = -\nabla\varphi. \quad (4.11)$$

The minus sign means that the vector \mathbf{E} is oriented in the direction of the most rapid decrease of the potential φ . The choice of such a direction is arbitrary.

The quantity

$$\int_1^2 \mathbf{E} \cdot d\mathbf{l} = - \int_1^2 \nabla\varphi \cdot d\mathbf{l} = \varphi_1 - \varphi_2, \quad (4.12)$$

called the electromotive force or, briefly, e.m.f., will occur often in what follows. (It should be noted that the electromotive force is not a force either in its nature or its dimensionality. The term e.m.f. has the vindication of historical tradition.) In the electrostatic field the electromotive force is equal to the difference between the electrostatic potentials at corresponding points.

Substituting the definition (4.11) into the equations of the electrostatic field (4.5) and (4.10), we see that (4.5) is satisfied identically, while (4.10) gives

$$\nabla \cdot \nabla\varphi = -4\pi\rho,$$

or, using formula (I. 49),

$$\nabla^2\varphi = -4\pi\rho. \quad (4.13)$$

This equation, called Poisson's equation, will be discussed in §14.

§5. The equation of continuity

Later on we shall have to pass over to the consideration of the more complicated case of the fields of moving charges. The motion of electric charges in space leads to the charge transport called electric current or, briefly, current. We characterize electric current by the current density vector $\mathbf{j}(\mathbf{r}, t)$, defined by the equality

$$\mathbf{j} = \sum e_i \mathbf{v}_i,$$

where e_i is the value of the charge, and \mathbf{v}_i is the velocity vector of the i th charge. The summation is carried out over all charges present at time t in a unit volume surrounding the point \mathbf{r} .

In the case of a continuous charge distribution the current density can be written in the form

$$\mathbf{j} = \rho \mathbf{v}. \quad (5.1)$$

The current density vector obviously represents the value of the charge perpendicular to the velocity per second crossing an imaginary unit area at the point \mathbf{r} at time t .

The values of the functions ρ and \mathbf{v} , i.e. charge density and the velocity of displacement of charge cannot be arbitrary but must satisfy the requirements of the charge conservation law.

Consider a closed surface, inside which there is a certain charge $e = \int \rho \, dV$, and find the derivative

$$-\frac{\partial}{\partial t} \int \rho \, dV.$$

Here the integration is carried out over the volume V enclosed within the surface S . The value of the derivative (taken with the minus sign) represents the decrease per unit time of the charge inside the surface S . Since electric charges do not vanish, and do not arise spontaneously, the rate of decrease of the charge in the volume V is equal to the flow of charge per second, out of the surface S enclosing this volume. Consequently, the equality

$$-\frac{\partial}{\partial t} \int \rho \, dV = \oint \rho \mathbf{v} \cdot d\mathbf{S} \quad (5.2)$$

holds. Passing over, in the last integral, to integration over the volume, we obtain

$$-\frac{\partial}{\partial t} \int \rho \, dV = \int \nabla \cdot (\rho \mathbf{v}) \, dV$$

On changing the order of the independent operations of integration and differentiation with respect to volume and time respectively, we have

$$-\int \frac{\partial \rho}{\partial t} \, dV = \int \nabla \cdot (\rho \mathbf{v}) \, dV$$

Because the integration volume is arbitrary, this equality gives

$$\left. \begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) &= 0, \\ \frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} &= 0. \end{aligned} \right\} \quad (5.3)$$

or

Formula (5.3), representing the mathematical expression of the charge conservation law, is called the equation of continuity.

For steady state processes, when the charge density distribution is not changing in time, the equation of continuity reads

$$\left. \begin{aligned} \nabla \cdot (\rho \mathbf{v}) &= 0, \\ \nabla \cdot \mathbf{j} &= 0. \end{aligned} \right\} \quad (5.4)$$

or

Equations (5.4) show that, for steady state processes, the current density vector is of solenoidal character. The trajectories of moving charges are closed, and the lines representing the vector \mathbf{j} form closed, non-intersecting, current tubes*.

Hence we shall make use of the notion of the total current I through a surface S . By definition

$$I = \int \mathbf{j} \cdot d\mathbf{S} = \int j_n dS,$$

where the integration is carried out with respect to the surface S . The current I gives the value of the total charge passing per second through the surface S .

§6. The electromagnetic field of charges moving with a constant velocity

We now go on to the study of the field of moving charges. We shall call this field the electromagnetic field. The properties of the electromagnetic field are essentially more complex than those of the electrostatic field. The establishment of the basic rules determining the behaviour of electromagnetic fields was in part the result of an experimental investigation of electromag-

* In the particular case of a system of spreading charges, current tubes are not closed, but go off at one end to infinity.

netic phenomena (Oersted, Ampère, Ohm and Faraday) and in part the result of a theoretical prediction (Maxwell) which was only later on confirmed experimentally (Hertz).

The exposition of the history of the development of electromagnetism is outside the scope of this book. However, it should be stressed that, since the atomic character of charge was discovered only in experiments at the end of the 19th and the beginning of the 20th century, all previous experiments and their theoretical interpretation referred to phenomena in material media. We shall present the results of these experiments in the language of microscopic physics which deals with charges moving in vacuum. In other words, without dwelling on the set-up of the experiments themselves, we shall present their results in a general form, in which the effect of the medium and actual conditions of the performance of the experiments are excluded.

The basic laws of the electromagnetic field, which will be set out below, at present rest not only on numerous and various experimental data but constitute the basis of contemporary electrical and radio technology.

We shall consider first of all the motion with a constant velocity of a set of charges along a tube or a line L (fig. I.1). In other words, we assume that an electric current, whose density \mathbf{j} does not depend on time, flows along the

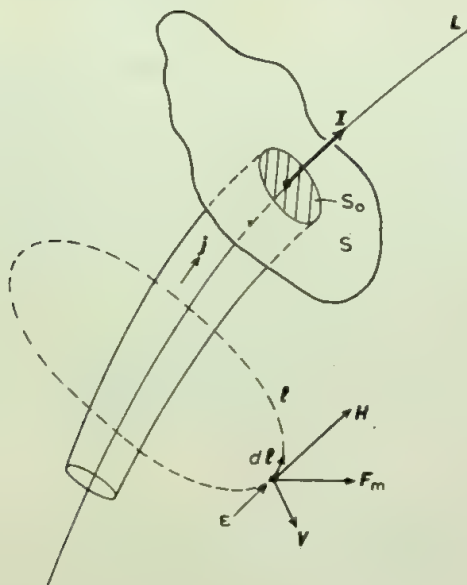


Fig. I.1

line. In practice an electric current can be realized most simply in metal conductors. However, since in this chapter we shall not deal with the motion of charges in material media, the current line will be understood to be not a metal conductor but a certain imaginary surface — a tube enclosing a set of curves — the trajectories of charged particles moving in vacuum.

We put a test charge in the vicinity of the path of the constant current. We are not interested in the effect of the charges moving along the path on the test charge e when at rest. If the test charge moves with respect to the current line with velocity v , new facts are observed. It turns out that a moving test charge allows one to detect a field which is inseparably associated with moving charges in the path of the current and whose character differs from that of the electrostatic field. This field is called the magnetic field. This term is connected with the fact that a similar field is produced by permanent magnets.

The magnetic field, like the electrostatic field, has a vector character. It is characterized by a certain vector \mathbf{H} , called the magnetic field strength or, briefly, the magnetic field.

Experiment shows that the test charge is acted upon by a force

$$\mathbf{F}_m = \frac{e}{c} \mathbf{v} \times \mathbf{H}. \quad (6.1)$$

This force is called the Lorentz force *.

As is seen from formula (6.1), the Lorentz force is perpendicular to the test charge velocity v and to the vector \mathbf{H} , and forms with them a right-hand screwed system.

The numerical factor of proportionality is determined experimentally, if it is required that the vector \mathbf{H} should have the same dimensions as the vector \mathbf{E} . In the CGSE system the numerical factor c is equal to $3 \times 10^{10} \text{ cm sec}^{-1}$, and is a universal constant numerically the same value as the velocity of light in vacuum.

As for any vector field, we introduce the basic characteristic of the magnetic field, the integral $\oint \mathbf{H} \cdot d\mathbf{l}$, called the magnetomotive force **.

A study of the magnetic fields of direct currents has shown that the value of the magnetomotive force is equal to

* We shall often call the Lorentz force the total force acting on a charged particle in an electric and magnetic field, equal to the sum of the forces caused by both fields.

** By analogy with the electromotive force in electrostatics. It should be stressed that the magnetomotive force, like the electromotive one, is a scalar and is called a force only by tradition.

$$\oint \mathbf{H} \cdot d\mathbf{l} = \frac{4\pi I}{c}, \quad (6.2)$$

where $I = \int \mathbf{j} \cdot d\mathbf{S}_0$ represents the total current per second passing through the cross section S_0 of the contour (tube) of the current.

Formula (6.2) shows that the magnetomotive force differs from zero only over a path which encloses the tube of current. In the simplest case of a rectilinear current (or a rectilinear segment of current), the vector lines of the magnetic field form a system of concentric circles about the current line (fig. 1.2). The value of the magnetomotive force is proportional to the total current I in the line. Formula (6.2) expresses the law of Oersted, who established in the years 1820–1826 the connection between the electric current and magnetic phenomena.

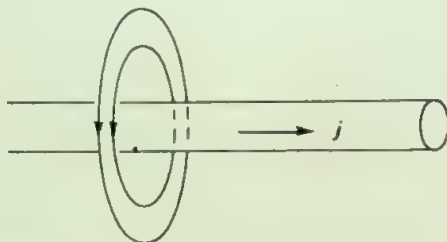


Fig. 1.2

As in electrostatics, experiment leads to an integral relation between the characteristics of the charge (the current I) and the field \mathbf{H} . In order to obtain a differential relation characterizing the field, we replace the integration surface S_0 in the expression for the total current by an arbitrary surface S drawn through the path of the current (fig. 1.1). The current density \mathbf{j} differs from zero only when the integration is carried out with respect to the cross section S_0 of the tube of current. Outside this cross section the current density is equal to zero, so that it can be written that

$$I = \int \mathbf{j} \cdot d\mathbf{S}_0 = \int \mathbf{j} \cdot d\mathbf{S}.$$

We then have

$$\oint \mathbf{H} \cdot d\mathbf{l} = \frac{4\pi}{c} \int \mathbf{j} \cdot d\mathbf{S}.$$

Making use of Stokes' theorem, we find

$$\int (\nabla \times \mathbf{H}) \cdot d\mathbf{S} = \frac{4\pi}{c} \int \mathbf{j} \cdot d\mathbf{S},$$

whence, in view of the arbitrariness of the surface S , it follows that

$$\nabla \times \mathbf{H} = \frac{4\pi}{c} \mathbf{j}. \quad (6.3)$$

We see that the magnetic field has vortex character. Eq. (6.3) determines the curl of the magnetic field at every point of space as a function of the value of the current density \mathbf{j} at every point of space.

Eq. (6.3) agrees with the stationary equation of continuity (5.4) and, hence, is self-consistent. Indeed, calculating the divergence of the two sides of (6.3) we arrive at the equality

$$\nabla \cdot (\nabla \times \mathbf{H}) = \frac{4\pi}{c} \nabla \cdot \mathbf{j} = 0. \quad (6.4)$$

Formula (6.3) shows that the motion of electric charges is inseparably associated with a magnetic field or, less rigorously, the motion of charges gives rise to a magnetic field.

In order to determine a magnetic field uniquely it is necessary to know the second differential relation $-\nabla \cdot \mathbf{H}$. To find it we shall consider the flux of the vector \mathbf{H} through an arbitrary closed surface $\oint \mathbf{H} \cdot d\mathbf{S}$. The experimental study of the distribution of the magnetic fields of direct currents shows that the magnetic fields are always of a purely solenoidal character, and

$$\oint \mathbf{H} \cdot d\mathbf{S} = 0. \quad (6.5)$$

Consequently,

$$\oint \mathbf{H} \cdot d\mathbf{S} = \int \nabla \cdot \mathbf{H} dV = 0$$

or, because of the arbitrariness of the volume of integration,

$$\nabla \cdot \mathbf{H} = 0. \quad (6.6)$$

Thus, the magnetic field has neither sources nor sinks. The lines of the magnetic field are always closed or go off to infinity *.

Eqs. (6.3) and (6.6) determine completely the magnetic field of direct currents. The magnetic field of direct currents possesses additive properties, as does the electrostatic field. This follows, in particular, from the linear character of the field eqs. (6.3) and (6.6). The magnetic and electrostatic fields are independent of each other. An electrostatic field (for a given distribution of moving charges) has no effect on the magnetic field of the charges.

In conclusion we note one more very important feature of the magnetic field \mathbf{H} . In contrast to the electrostatic field \mathbf{E} , which is characterized by a polar vector, the vector of the strength of the magnetic field \mathbf{H} is an axial vector or a pseudo-vector. This is seen from the formula for the Lorentz force (6.1). Indeed, from the definition of the field strength by formula (6.1) it is seen that the behaviour of the vector \mathbf{H} in the reflection $\mathbf{r} \rightarrow (-\mathbf{r})$ through the origin is determined by the behaviour of the polar vectors \mathbf{F} and \mathbf{v} and the properties of their vector product. In substituting $\mathbf{r} \rightarrow (-\mathbf{r})$ the directions of the vectors \mathbf{F} and \mathbf{v} become reversed, and the sign of the vector product is changed. Consequently, in substituting $\mathbf{r} \rightarrow (-\mathbf{r})$ the vector \mathbf{H} must remain unchanged. It is just this property that is an indication of an axial vector.

§ 7. The electromagnetic field of moving charges. The general case

In considering the non-steady state motion of charges or, what is the same, the non-steady currents in certain contours, further results are obtained.

It should be stressed, first of all, that eq. (6.3) cannot remain valid for non-steady currents. As we have just seen, eq. (6.3) leads to the charge conservation law in the form of (5.4). However, for non-steady state processes the law is expressed by formula (5.3). Thus, for these processes the relation (6.3) contradicts the charge conservation law.

The most important fact which in essence distinguishes non-stationary magnetic and electric fields from stationary ones is the existence of an interrelation between them.

Faraday observed that a change of a magnetic field in time entails the appearance of an electric field (the phenomenon of electromagnetic induc-

* Strictly speaking, for complex configurations of currents there may also exist lines of field which are not closed but fill a surface densely. See: I. E. Tamm, *Osnovy teorii elektrichestva (Introduction to the theory of electricity)* (Gostekhizdat, 1954) p. 53.

tion). Maxwell predicted theoretically that a change of an electric field in time leads to the appearance of a magnetic field. This theoretical prediction subsequently found confirmation in Hertz' experiments.

Faraday's experiments established that a change in time of the flux of a magnetic field vector through an arbitrary surface S is accompanied by the appearance of an electromotive force in the contour enclosing this surface, i.e.

$$\oint \mathbf{E} \cdot d\mathbf{l} = - \frac{1}{c} \frac{\partial}{\partial t} \int \mathbf{H} \cdot d\mathbf{S}. \quad (7.1)$$

The factor of proportionality c turned out to be numerically equal to $3 \times 10^{10} \text{ cm sec}^{-1}$, i.e. to the velocity of light in vacuum. Fig. 3 shows schematically the relation between the change in the magnetic field $\partial \mathbf{H} / \partial t$ and the electromotive force. If the lines of the field vector $\partial \mathbf{H} / \partial t$ are represented by straight lines going from the left to the right, then the lines of the electric field are represented by concentric circles embracing the corresponding straight lines. The directions of the vectors of the electric field \mathbf{E} and $\partial \mathbf{H} / \partial t$ are shown in fig. 1.3 by arrows.

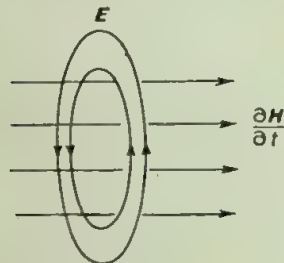


Fig. 1.3

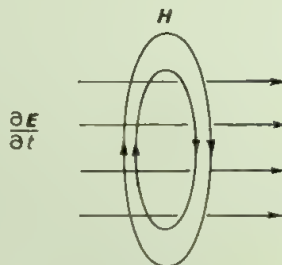


Fig. 1.4

Formula (7.1) represents the generalized Faraday law of induction (1831). The generalization consists in the following. The experimental data of Faraday referred to a circuit of a metal wire. The appearance of an induced electric field in a conductor corresponds to the appearance of a current in the circuit. It was this current that was measured directly. In formula (7.1) the integration is carried out over a completely arbitrary contour independent of the presence of conductors. This means that the primary factor is the appearance of the field in the contour. The electric current is some secondary phenomenon associated with the material nature of the contour — the presence of a conductivity in it. The generalized Faraday law of induction in the

form (7.1) establishes the interrelation between the magnetic and electric field. The sign in formula (7.1) corresponds to the well-known Lenz rule.

Maxwell put forward the hypothesis, subsequently confirmed experimentally, that in addition to the law (7.1) there is an analogous relation between the change of an electric field in time and the magnetic field.

Namely, when the flux of an electric field through a surface S changes in time,

$$\frac{\partial}{\partial t} \int \mathbf{E} \cdot d\mathbf{S},$$

there arises in the contour embracing the surface S a magnetomotive force equal to

$$\oint \mathbf{H} \cdot d\mathbf{l} = \frac{1}{c} \frac{\partial}{\partial t} \int \mathbf{E} \cdot d\mathbf{S}, \quad (7.2)$$

where the coefficient c has the same value as in formula (7.1). Fig. 1.4 illustrates the relation between the change of the electric field, characterized by the vector $\partial \mathbf{E} / \partial t$, and the lines of the magnetic field. If the vector $\partial \mathbf{E} / \partial t$ at every point of the field is represented by a family of straight lines, then the lines of the magnetic field are represented by concentric circles about these straight lines.

We shall postpone till §9 the discussion of the considerations which lead to the establishment of the relation between the change of the electric field and the magnetic field circulation. As to the connection of the law (7.2) with experimental data, it turns out that the existence of electromagnetic waves is associated with (7.2). If the line L , along which the magnetic field circulation is calculated, links the contour defining the electric current, then the total magnetomotive force is expressed by the formula (see (6.2) and (7.2))

$$\oint \mathbf{H} \cdot d\mathbf{l} = \frac{4\pi}{c} \int \mathbf{j} \cdot d\mathbf{S} + \frac{1}{c} \frac{\partial}{\partial t} \int \mathbf{E} \cdot d\mathbf{S}. \quad (7.3)$$

From formulae (7.1) and (7.3), containing the integral characteristics of the fields \mathbf{E} and \mathbf{H} , one can pass over to differential characteristics. By means of Stokes' theorem formula (7.1) can be rewritten in the form

$$\int (\nabla \times \mathbf{E}) \cdot d\mathbf{S} = -\frac{1}{c} \frac{\partial}{\partial t} \int \mathbf{H} \cdot d\mathbf{S} = -\frac{1}{c} \int \frac{\partial \mathbf{H}}{\partial t} \cdot d\mathbf{S}. \quad (7.4)$$

We have changed the order of carrying out the independent operations of differentiation with respect to time and integration with respect to a surface fixed in space (see, however, Part IV, §23).

In view of the arbitrariness of the integration surface in formula (7.4), there follows from the equality of the integrals the equality of the integrands:

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t}. \quad (7.5)$$

We see that a time varying electric field is, in general, rotational, in contrast to the electrostatic field. The value of the curl of the vector \mathbf{E} at a given point is determined by the rate of change in time of the magnetic field at the same point.

Transforming the integral relation (7.3) in a completely analogous way, we have

$$\nabla \times \mathbf{H} = \frac{4\pi}{c} \mathbf{j} + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} \quad (7.6)$$

This relation shows that the curl of the magnetic field at every point of space is determined by the current density of the charges \mathbf{j} and the change of the electric field in time.

The symmetry of the equations for the curls of the electric and magnetic fields (7.5) and (7.6) is striking. A change in the magnetic field is accompanied by the appearance of a rotational electric field. Conversely, a change in the electric field is associated with the appearance of a rotational magnetic field. However, besides the similarity between these equations there is also an extremely important difference between them. First, eqs. (7.5) and (7.6) differ from each other by the sign before the time derivatives. Second, in general the curl of the magnetic field depends not only on the rate of change of the electric field but also on the current density \mathbf{j} . However, there are no 'magnetic charges' which can move in space and produce a 'current of magnetic charges'.

Eqs. (7.3) and (7.5) determine the curls of the electric and magnetic fields. For a single-valued determination of the fields there must also be given the divergences of the electric and magnetic fields.

In order to determine these we shall find the divergences of both sides of eqs. (7.5) and (7.6). We start from the latter equation. We have

$$\nabla \cdot (\nabla \times \mathbf{H}) = 0 = \frac{4\pi}{c} \nabla \cdot \mathbf{j} + \frac{1}{c} \nabla \cdot \frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c} \nabla \cdot \mathbf{j} + \frac{1}{c} \left(\frac{\partial}{\partial t} \nabla \cdot \mathbf{E} \right).$$

By means of the charge conservation law (5.3) the above equation can easily be transformed into

$$\frac{\partial}{\partial t} (\nabla \cdot \mathbf{E} - 4\pi\rho) = 0,$$

whence

$$\nabla \cdot \mathbf{E} = 4\pi\rho + f(\mathbf{r}),$$

where the function $f(\mathbf{r})$ is any function depending only on coordinates but not on time.

Let us assume that at an arbitrary initial instant of time, the charges were at rest. Then at the initial instant the electrostatic field equation (4.10) must have been satisfied, and the function $f(\mathbf{r})$ was equal to zero. Since $f(\mathbf{r})$ does not depend on time, this means that $f(\mathbf{r})$ is always equal to zero. The electric field divergence for stationary as well as non-stationary fields is determined by the formula

$$\nabla \cdot \mathbf{E} = 4\pi\rho. \quad (7.7)$$

Similarly, taking the divergence of (7.5), we find

$$\nabla \cdot (\nabla \times \mathbf{E}) = 0 = \nabla \cdot \left[\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t} \right]$$

or

$$\nabla \cdot \mathbf{H} = f_1(\mathbf{r}).$$

Assuming again that at the initial moment the magnetic field had a stationary character, and repeating the reasoning just presented, we arrive at the conclusion

$$\nabla \cdot \mathbf{H} = 0 \quad (7.8)$$

The magnetic field for stationary as well as non-stationary fields has a purely rotational (solenoidal) character. Neither in stationary nor in variable fields are there magnetic charges on which the lines of the magnetic field begin and end.

§8. Maxwell-Lorentz system of equations

The system of eqs. (7.5)–(7.8) which we have obtained for the electromagnetic field in a vacuum is called Maxwell's equations. They were established by Maxwell in 1873 for the more general case of electromagnetic fields in material media, and by Lorentz in 1895 for a system of charges moving in vacuum.

Let us write once again Maxwell's equations, grouping them into two pairs:

	Differential form	Integral form	
Pair I	$\begin{cases} \nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t}, \\ \nabla \cdot \mathbf{H} = 0, \end{cases}$	$\begin{cases} \oint \mathbf{E} \cdot d\mathbf{l} = -\frac{1}{c} \frac{\partial}{\partial t} \int \mathbf{H} \cdot d\mathbf{S}; \\ \oint \mathbf{H} \cdot d\mathbf{S} = 0 \end{cases}$	(8.1) (8.2)

Pair II	$\begin{cases} \nabla \times \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi \mathbf{j}}{c}, \\ \nabla \cdot \mathbf{E} = 4\pi \rho \end{cases}$	$\begin{cases} \oint \mathbf{H} \cdot d\mathbf{l} = \frac{4\pi}{c} I + \frac{1}{c} \frac{\partial}{\partial t} \int \mathbf{E} \cdot d\mathbf{S}; \\ \oint \mathbf{E} \cdot d\mathbf{S} = 4\pi e. \end{cases}$	(8.3) (8.4)
---------	---	--	----------------

Assuming that the distributions of the currents and charges are known, one can find by means of Maxwell's equations the six unknown components of the field vectors \mathbf{E} and \mathbf{H} . As we have seen in the preceding paragraph, the equations for the divergence of \mathbf{E} and \mathbf{H} follow from those for the curl and the initial conditions. Hence in the Maxwell system of eight scalar differential equations there are only six independent equations.

Eq. (8.1), representing a generalization of the Faraday induction law, establishes that a magnetic field changing with time gives rise to a rotational electric field. Eq. (8.2) shows that the magnetic field has a solenoidal character and that the lines of the magnetic field are either closed or go off to infinity.

From eq. (8.3) it follows that a rotational magnetic field is produced when charges are moving and when the electric field changes with time. By analogy with the electric current, the quantity $(1/4\pi)(\partial \mathbf{E}/\partial t)$ in the first term on the right is called the *displacement current*, while the sum of the two terms is called the *total current*. It can then be said that the rotational magnetic field is produced by the total current into which the two terms enter equivalently. Finally, eq. (8.4) shows that the sources of the electric field are electric charges. For given density distribution of charges and currents Maxwell's

equations determine completely the electric field $\mathbf{E}(\mathbf{r}, t)$ and the magnetic field $\mathbf{H}(\mathbf{r}, t)$.

Maxwell's equations represent a system of linear partial differential equations. By virtue of the linearity of the equations, the principle of superposition of electromagnetic fields holds. If \mathbf{E}_i and \mathbf{H}_i are solutions of the Maxwell equations, then $\mathbf{E} = \sum \mathbf{E}_i$ and $\mathbf{H} = \sum \mathbf{H}_i$ are also solutions of these equations.

The integration of partial differential equations becomes definite only in the case where a set of boundary and initial conditions is given. We shall discuss boundary and initial conditions later, in § 24.

Up to now we have said nothing about the distribution of the charges and their motion in space. However, the distribution of the charges and their velocities cannot be given quite arbitrarily. The charge density and the current density (the velocities of the charges) are interrelated by the charge conservation law:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0. \quad (8.5)$$

Charges moving in an electromagnetic field are acted upon by the Lorentz force. The equations of motion of a charge can be written in the form

$$\frac{d\mathbf{p}}{dt} = e \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right), \quad (8.6)$$

where \mathbf{p} is the momentum of the particle.

The field which must be substituted into (8.6) represents the total field including that produced by the charge itself as well as the external field produced by other charges. The former must also have an effect on the motion of the particle. However, in most cases one can assume the self field to be weak and not take it into account (see § 29). In this approximation the vectors \mathbf{E} and \mathbf{H} in (8.6) stand for the external field acting on the particle. The law of motion (8.6) can also be written for continuously distributed charges, if \mathbf{p}_0 is understood to be the momentum of the particles in unit volume and if the force density (i.e. the force per unit volume) is substituted for the Lorentz force. Then

$$\frac{d}{dt} \int \mathbf{p}_0 dV = \int \rho \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right) dV, \quad (8.7)$$

or

$$\frac{d}{dt} \mathbf{p}_0 = \rho \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right). \quad (8.8)$$

We have already mentioned that field equations were initially formulated by Maxwell for electromagnetic processes in matter. Lorentz established their applicability to a system consisting of a field and charges, and supplemented them by the equation of motion of the charges.

Hence the eqs. (8.1)–(8.8) are often called the Maxwell-Lorentz equations. The Maxwell-Lorentz equations contain the complete description of the behaviour of a system consisting of fields and charges. If the values of the functions ρ and \mathbf{v} and the initial values of the fields \mathbf{E} and \mathbf{H} are given, then the integration of these equations allows one to find the electric and magnetic field distribution in space at any subsequent instant of time. Thus, in electrodynamics, as in mechanics, if the state of a system at the initial moment and the law of change of state are given one can determine uniquely the state at a subsequent moment.

It should be noted that, within the framework of this chapter, the Lorentz force acting on a moving charge should be assumed as an empirical formula. In Chapter II it will be shown that the expression for the Lorentz force follows, as a consequence, from more general laws of physics.

The range of applicability of the Maxwell-Lorentz equations is extremely large. They determine the character of electromagnetic processes on the cosmic scale, constitute the basis of contemporary electrical and radio technology, and allow one to investigate electromagnetic phenomena taking place with individual charges. But, nevertheless, as we shall see in what follows, the Maxwell-Lorentz equations and the classical field theory based on them are not an expression of the universal laws of nature, and have a restricted region of applicability.

A number of electromagnetic processes and, above all, intra-atomic ones, lie beyond the limits of applicability of the Maxwell-Lorentz equations. The problem of establishment of these limits will be discussed more than once later on.

§9. The displacement current

In contrast with the electric current \mathbf{j} , which has a very simple and obvious meaning, the displacement current $(1/4\pi)(\partial\mathbf{E}/\partial t)$ is not associated with the motion of any charges.

The displacement current was introduced by Maxwell who interpreted it in terms of ether theory, which was generally accepted at the time but which has now been abandoned as erroneous.

It is readily understood why, in the middle of the 19th century, one could not have had recourse to experimental data in order to obtain formula (7.6). The velocities of motion of electric charges are, as a rule, very large. Hence the current ρv is always large in comparison with the displacement current $(1/4\pi)(\partial E/\partial t)$, provided the electric field is not changing very rapidly in time. Estimates show that the two terms of the total current may be of the same order of magnitude for periodic variation of the vector E with a frequency of the order of $10^6 - 10^7 \text{ sec}^{-1}$ or higher.

This is a range of radio frequencies which was unknown in physics in the middle of the 19th century. It was only in 1888 that Hertz was the first to establish experimentally the existence of electromagnetic waves, and hence prove the reality of the displacement current.

The fact that the current j cannot always be responsible for the appearance of a magnetic field is very clearly illustrated by the following reasoning. Let an electric charge e be moving in space. We seek to determine the strength of the magnetic field produced by this charge along a certain contour L (fig. 1.5). According to Stokes' theorem,

$$\oint \mathbf{H} \cdot d\mathbf{l} = \int (\nabla \times \mathbf{H}) \cdot d\mathbf{S} = \frac{4\pi}{c} \int \mathbf{j} \cdot d\mathbf{S},$$

where the integration surface is any surface enclosed by the contour L .

Let us consider at a certain instant two surfaces S_1 and S_2 enclosed by the contour L . It is clear that from the standpoint of the electric current passing through them they are not equivalent. No current passes through the surface S_1 at the given instant because the charge e has not yet reached it. On the

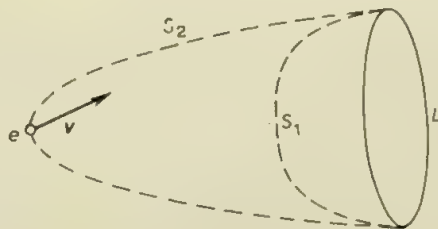


Fig. 1.5

contrary, the surface S_2 is crossed by the charge and the current through it differs from zero. Thus, we arrive at the obvious contradiction:

$$\int_{S_1} (\nabla \times \mathbf{H}) \cdot d\mathbf{S}_1 = 0 \quad \text{and} \quad \int_{S_2} (\nabla \times \mathbf{H}) \cdot d\mathbf{S}_2 \neq 0,$$

although according to Stokes' theorem these integrals must be equal to each other. In reality the moving charge produces at the surface S_1 a certain electric field changing in time. A calculation which will be performed in §20 shows that the derivative of this electric field with respect to time, integrated with respect to the surface S_1 , gives exactly the same value for $\oint \mathbf{H} \cdot d\mathbf{l}$ in the contour L as the integral $4\pi \int \mathbf{j} \cdot d\mathbf{S}$ with respect to the surface S_2 . Thus, it is necessary to assume equivalence (in the sense of producing circulation of the magnetic field in a certain contour) of the charge current and the change in the electric field for any moving charge.

The introduction of displacement current can be made on the basis of the following formal reasoning: it is necessary to find a generalization of the law (6.3) such that it does not contradict the charge conservation law (5.3) for a non-stationary change of the field.

Writing (6.3) in the form

$$\nabla \times \mathbf{H} = \frac{4\pi}{c} (\mathbf{j} + \mathbf{C}), \quad (9.1)$$

where \mathbf{C} is an unknown vector, and taking the divergence of both sides of (9.1), we find

$$\nabla \cdot (\nabla \times \mathbf{H}) = 0 = \frac{4\pi}{c} (\nabla \cdot \mathbf{j} + \nabla \cdot \mathbf{C}),$$

or

$$\nabla \cdot (\mathbf{j} + \mathbf{C}) = 0. \quad (9.2)$$

The unknown vector \mathbf{C} supplements the current density \mathbf{j} in such a way that the quantity $(\mathbf{j} + \mathbf{C})$ possesses the properties of a solenoidal vector having closed current lines.

The divergence of the vector \mathbf{C} can be found from eqs. (9.2), (5.3) and (4.10). We have

$$\nabla \cdot \mathbf{C} = -\nabla \cdot \mathbf{j} = \frac{\partial \rho}{\partial t} = \frac{1}{4\pi} \frac{\partial}{\partial t} \nabla \cdot \mathbf{E} = \frac{1}{4\pi} \nabla \cdot \frac{\partial \mathbf{E}}{\partial t},$$

whence it follows that the vector \mathbf{C} is equal to

$$\mathbf{C} = \frac{1}{4\pi} \frac{\partial \mathbf{E}}{\partial t} + \mathbf{b}(\mathbf{r}, t),$$

where $\mathbf{b}(\mathbf{r}, t)$ is any vector satisfying the condition

$$\nabla \cdot \mathbf{b}(\mathbf{r}, t) = 0.$$

There are no grounds to assume beforehand that the vector \mathbf{b} is equal to zero.

The brilliant idea of Maxwell consisted in his assumption that a profound symmetry and interrelation exist between the electric and magnetic fields. We emphasized this symmetry in considering the properties of Maxwell's equations. To obtain symmetry between the fields it is necessary to assume $\mathbf{b} \equiv 0$. Then

$$\nabla \times \mathbf{H} = \frac{4\pi}{c} \mathbf{j} + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} \quad (9.3)$$

and, in particular, in the space region in which there are no moving charges $\mathbf{j} = 0$ and the equation for $\nabla \times \mathbf{H}$ is of the form

$$\nabla \times \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t}$$

The curl of the vector \mathbf{H} is determined by the rate of change of the vector \mathbf{E} , whereas, according to Faraday's law of induction (7.5), the curl of the vector \mathbf{E} is determined by the rate of change of the vector \mathbf{H} . Only in those regions of space where the current density differs from zero is the symmetry between electric and magnetic field violated. Since there are no real magnetic charges and since the magnetic field is always of purely solenoidal character, then in the equation for the electromagnetic induction (7.5) there is no term analogous to the $(4\pi/c)\mathbf{j}$ of eq. (9.3). On the contrary, if the vector \mathbf{b} were different from zero, no symmetry would exist between the electric and magnetic fields.

In conclusion we emphasize that since the displacement current in vacuum $(1/4\pi)(\partial \mathbf{E}/\partial t)$ is not connected with displacement or change of state of any particles, it cannot be compared with any mechanical model allowing one to picture this physical quantity in a simple way. For the sake of clarity, particles at rest or in motion — e.g. small spheres or pellets — were associated

with the ideas of "charge density" or "current density". The field vectors \mathbf{E} and \mathbf{H} were represented by lines of force or field lines, which were previously pictured in the form of tensions in an elastic medium — the electromagnetic ether. Such mechanical models were undoubtedly useful, because they helped one to understand clearly the meaning of the corresponding quantities. Displacement current in a vacuum is the first quantity we encounter which cannot be described by means of a mechanical analogue. In what follows we shall have to deal with a very large number of important physical notions and quantities having no mechanical analogues. Like the displacement current in vacuum, they cannot be represented by any obvious model.

§ 10. The electromagnetic field potentials

In § 2 we have seen that it is convenient to introduce auxiliary quantities defined by relations (2.7) and (2.13), in order to find the stationary vector field from the values of divergence and curl given in each point of space. It turns out that one can proceed in exactly the same way in the more general case of a non-stationary system of vector fields — electric and magnetic.

The magnetic field vector is always solenoidal and its divergence is equal to zero. Hence, as in § 2, we assume that

$$\mathbf{H} = \nabla \times \mathbf{A}, \quad (10.1)$$

where the auxiliary vector \mathbf{A} is called the vector potential. The vector potential $\mathbf{A}(\mathbf{r}, t)$ is a function of space and time. The eq. (8.2) is automatically satisfied, since for any $\mathbf{A}(\mathbf{r}, t)$

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0$$

Substitution of the relation (10.1) into (8.1) gives

$$\nabla \times \mathbf{E} = -\frac{1}{c} \nabla \times \frac{\partial \mathbf{A}}{\partial t}$$

or

$$\nabla \times \left(\mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \right) = 0.$$

The last equation shows that the vector

$$\mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t}$$

is a potential vector, i.e. it can be written in the form

$$\mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} = -\nabla \varphi,$$

where φ is a function of coordinates and time which we shall call the scalar potential.

In contrast with the electrostatic case, the electric field vector, having a vortex nature, cannot be written in the form of the gradient of a potential. It is expressed through the totality of scalar and vector potentials by the formula

$$\mathbf{E} = -\nabla \varphi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t}, \quad (10.2)$$

where the second term, connecting the electric field with magnetic quantities, expresses the law of electromagnetic induction.

For the determination of \mathbf{A} and φ we have in addition eqs. (8.3) and (8.4). The former equation gives

$$\nabla \times (\nabla \times \mathbf{A}) = -\frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \frac{1}{c} \nabla \frac{\partial \varphi}{\partial t} + \frac{4\pi}{c} \mathbf{j},$$

or, according to formula (I. 50),

$$\nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = -\frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \frac{1}{c} \nabla \frac{\partial \varphi}{\partial t} + \frac{4\pi}{c} \mathbf{j}.$$

We rewrite the last equation in the form

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\frac{4\pi}{c} \mathbf{j} + \nabla \left(\nabla \cdot \mathbf{A} + \frac{1}{c} \frac{\partial \varphi}{\partial t} \right). \quad (10.3)$$

Eq. (8.4) gives in turn

$$\nabla^2 \varphi = -4\pi\rho - \frac{1}{c} \frac{\partial}{\partial t} \nabla \cdot \mathbf{A}. \quad (10.4)$$

The potentials \mathbf{A} and φ are auxiliary quantities, introduced in order to simplify the field equations. We shall impose upon them those conditions which will allow us to make eqs. (10.3) and (10.4) independent without changing relations (10.1) and (10.2). Relation (10.1) determines the curl of the vector potential \mathbf{A} . However, the vector potential \mathbf{A} is in itself still undetermined, since its divergence is not given. If the divergence of \mathbf{A} is given by the relation

$$\nabla \cdot \mathbf{A} + \frac{1}{c} \frac{\partial \varphi}{\partial t} = 0, \quad (10.5)$$

called the Lorentz relation, then (10.3) will turn into

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\frac{4\pi}{c} \mathbf{j}. \quad (10.6)$$

Eq. (10.4) will assume the completely analogous form

$$\nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = -4\pi \rho. \quad (10.7)$$

The equations obtained for the electromagnetic potentials are equivalent to Maxwell's equations. If the distributions of the charge $\rho(\mathbf{r}, t)$ and the currents $\mathbf{j}(\mathbf{r}, t)$ (satisfying the continuity equation) are given, then integration of eqs. (10.6) and (10.7) allows us to find the vector potential and scalar potential. The field vectors will be found by differentiating according to formulae (10.1) and (10.2).

An equation of the type

$$\nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = \psi(\mathbf{r}, t),$$

where $\psi(\mathbf{r}, t)$ is a function of coordinates and time, is called D'Alembert's equation. It is often written by means of the so-called D'Alembert differential operator (D'Alembertian):

$$\square = \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}$$

in the more compact form:

$$\square\varphi = \psi(\mathbf{r}, t) .$$

In the particular case of the homogeneous equation for which $\psi(\mathbf{r}, t) = 0$, the so-called wave equation,

$$\nabla^2\varphi - \frac{1}{c^2} \frac{\partial^2\varphi}{\partial t^2} = 0 ,$$

is obtained from D'Alembert's equation. In another particular case, that of time-independent functions $\varphi(\mathbf{r})$ and $\psi(\mathbf{r})$, D'Alembert's equation reduces to Poisson's equation which is already known (see (2.8)) from electrostatics:

$$\nabla^2\varphi = \psi(\mathbf{r}) .$$

From the mathematical standpoint the second-order equations — Poisson's equation, the wave equation and D'Alembert's equation — are simpler than Maxwell's first-order equations in partial derivatives. As we shall see in what follows, the general solution of D'Alembert's equation and the wave equation can be obtained in integral form, as for Poisson's equation (see (2.9)). It is for this reason that, in investigating the properties of electromagnetic fields, as well as in solving a number of concrete problems, the use of potentials is very convenient and the potential method represents the basic mathematical apparatus of the field theory.

§ 11. Gauge invariance of the potentials

We have already emphasized that the potentials \mathbf{A} and φ represent auxiliary quantities having no direct physical meaning. The field strengths \mathbf{E} and \mathbf{H} , which have a definite value at every point of the field and at any instant, have a real meaning. They can be measured by means of the forces acting on a test charge moving in the field.

The values of \mathbf{A} and φ cannot be measured and, therefore, the potentials by themselves should not be involved in the final expressions of the field theory. Indeed, determination of the potentials \mathbf{A} and φ from formulae (10.1) and (10.2) is not unique and allows a certain arbitrariness.

We now have to discuss the question of the degree of arbitrariness in the determination of the potentials in the general case. From the definition of the vector potential (10.1) it follows that if we perform the transformation

$$\mathbf{A} \rightarrow \mathbf{A}' + \nabla \psi(\mathbf{r}, t), \quad (11.1)$$

where $\psi(\mathbf{r}, t)$ is an arbitrary function of coordinates and time, then we arrive at the same value of the field strength \mathbf{H} :

$$\mathbf{H} = \nabla \times \mathbf{A} = \nabla \times (\mathbf{A}' + \nabla \psi(\mathbf{r}, t)) = \nabla \times \mathbf{A}' + \nabla \times \nabla \psi = \nabla \times \mathbf{A}'.$$

Now consider the determination of the scalar potential φ given by formula (10.2). The transformation (11.1) leads to the value

$$\mathbf{E} = -\nabla \varphi - \frac{1}{c} \frac{\partial \mathbf{A}'}{\partial t} - \frac{1}{c} \nabla \frac{\partial \psi}{\partial t} = -\nabla \left(\varphi + \frac{1}{c} \frac{\partial \psi}{\partial t} \right) - \frac{1}{c} \frac{\partial \mathbf{A}'}{\partial t}.$$

Substituting

$$\varphi \rightarrow \varphi' - \frac{1}{c} \frac{\partial \psi}{\partial t}, \quad (11.2)$$

we arrive at the previous expression for the electric field strength.

Thus, the vector potential is determined to within a vector representing the gradient of an arbitrary function of coordinates and time $\psi(\mathbf{r}, t)$, while the scalar potential is determined to within the derivative of the same function with respect to time. In particular, in the case of a time-independent electric field, one can add an arbitrary constant to the potential φ and arrive at the same value of the field strength $\mathbf{E} = -\nabla(\varphi + \text{const}) = -\nabla \varphi$.

In general it can be said that two fields described by the system of potentials \mathbf{A}' and φ' and \mathbf{A} and φ respectively, are physically identical if \mathbf{A} and \mathbf{A}' , φ and φ' can be connected with each other by the relations (11.1) and (11.2). The same fact can be expressed by the words: electromagnetic field equations are invariant under the transformations (11.1) and (11.2).

Different ways of choosing the potentials \mathbf{A} and φ , leaving the field strengths \mathbf{E} and \mathbf{H} unchanged, are called different gauges of the potentials. The invariance of the fields \mathbf{E} and \mathbf{H} and all other relations of the field theory with respect to different gauges are called gauge invariance. The property of gauge invariance permits one, by allowing a certain degree of arbitrariness in choosing electromagnetic potentials, to select them in such a way that the relations of the field theory may take the simplest form.

Lorentz's condition, introduced in the previous paragraph, serves as an example of such a selection.

We can now show that Lorentz's condition corresponds to a definite gauge

of potentials (Lorentz gauge). Let the Lorentz condition not be fulfilled for certain A_0 and φ_0 , so that

$$\nabla \cdot A_0 + \frac{1}{c} \frac{\partial \varphi_0}{\partial t} = \chi(\mathbf{r}, t) \neq 0.$$

We perform the gauge transformation (11.1) and (11.2):

$$A_0 \rightarrow A + \nabla \psi,$$

$$\varphi_0 \rightarrow \varphi - \frac{1}{c} \frac{\partial \psi}{\partial t}.$$

Then we have

$$\nabla \cdot A + \frac{1}{c} \frac{\partial \varphi}{\partial t} = \chi(\mathbf{r}, t) - \nabla^2 \psi + \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2}. \quad (11.3)$$

If it is required that the function ψ should satisfy D'Alembert's equation

$$\nabla^2 \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = \chi(\mathbf{r}, t), \quad (11.4)$$

then the Lorentz condition will be fulfilled for the transformed potentials A and φ . We have already mentioned that D'Alembert's equation always has a solution. Hence, for a given value of χ it is always possible to select a function ψ satisfying (11.4).

It should be stressed that the arbitrary function $\psi(\mathbf{r}, t)$ is not determined completely by eq. (11.4). One can add to it a function $\psi_0(\mathbf{r}, t)$ representing the solution of the homogeneous equation

$$\nabla^2 \psi_0 - \frac{1}{c^2} \frac{\partial^2 \psi_0}{\partial t^2} = 0.$$

Performing the transformations

$$A \rightarrow A' + \nabla \psi_0,$$

$$\varphi \rightarrow \varphi' - \frac{1}{c} \frac{\partial \psi_0}{\partial t},$$

we arrive again at the same values of field strengths \mathbf{E} and \mathbf{H} :

$$\mathbf{H} = \nabla \times \mathbf{A}',$$

$$\mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{A}'}{\partial t} - \nabla \varphi',$$

and the potentials \mathbf{A}' and φ' satisfy the Lorentz condition:

$$\nabla \cdot \mathbf{A}' + \frac{1}{c} \frac{\partial \varphi'}{\partial t} = 0.$$

Therefore, remaining within the framework of the Lorentz gauge, one can choose the function ψ in such a way that an additional condition, imposed upon one of the four values φ, A_x, A_y, A_z , can be fulfilled.

In addition to the Lorentz gauge use is sometimes made (particularly in quantum field theory) of another gauge, the so-called Coulomb gauge, for which

$$\nabla \cdot \mathbf{A} = 0.$$

In this gauge the equations for potentials (eqs. (10.3) and (10.4)) assume the form

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \frac{1}{c} \nabla \frac{\partial \varphi}{\partial t} = -\frac{4\pi}{c} \mathbf{j},$$

$$\nabla^2 \varphi = -4\pi\rho,$$

and the scalar potential φ is determined by the distribution of charges as if they were at rest. It goes without saying that the field strengths \mathbf{E} and \mathbf{H} found from the solutions of the equations for the potentials with the Coulomb gauge and with the Lorentz gauge are the same. In this book we shall, as a rule, make use of the Lorentz gauge.

§ 12. Energy conservation law of the electromagnetic field

The first important general consequence following from Maxwell's equations is the existence of the electromagnetic field energy. In order to find this

energy we shall consider a closed system consisting of a field and particles. Let us find the work W done by the field forces on the particles in a volume V . By considering the rate at which this is done, assuming the charges to be continuously distributed in space, and making use of (8.8), we can write that

$$\begin{aligned}
 \frac{dW}{dt} &= \int \mathbf{F} \cdot \mathbf{v} dV = \int \rho \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right) \cdot \mathbf{v} dV = \\
 &= \int \rho \mathbf{E} \cdot \mathbf{v} dV + \frac{1}{c} \int \rho (\mathbf{v} \times \mathbf{H}) \cdot \mathbf{v} dV = \\
 &= \int \mathbf{j} \cdot \mathbf{E} dV + \frac{1}{c} \int \rho (\mathbf{v} \times \mathbf{v}) \cdot \mathbf{H} dV = \int \mathbf{j} \cdot \mathbf{E} dV.
 \end{aligned} \tag{12.1}$$

The work of the magnetic field force is equal to zero, since this force is perpendicular to the velocity of the particle.

We transform the relation (12.1), making use of one of Maxwell's equations. Expressing the current density in terms of the field vectors by means of (8.3), we have

$$\frac{dW}{dt} = \int \mathbf{j} \cdot \mathbf{E} dV = \frac{c}{4\pi} \int \mathbf{E} \cdot (\nabla \times \mathbf{H}) dV - \frac{1}{4\pi} \int \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} dV. \tag{12.2}$$

As we have stressed more than once, there must exist a symmetry between the electric and magnetic fields. However, eq. (12.2) is asymmetric. We can make it have the symmetric form by adding to its right side the expression

$$- \frac{c}{4\pi} \int \mathbf{H} \cdot \left(\nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{H}}{\partial t} \right) dV,$$

which, from the Maxwell eq. (8.1), is equal to zero. This gives

$$\begin{aligned}
 \frac{dW}{dt} &= \int \mathbf{j} \cdot \mathbf{E} dV = \\
 &= \frac{c}{4\pi} \int [\mathbf{E} \cdot (\nabla \times \mathbf{H}) - \mathbf{H} \cdot (\nabla \times \mathbf{E})] dV - \int \frac{\partial}{\partial t} \left(\frac{E^2 + H^2}{8\pi} \right) dV
 \end{aligned} \tag{12.3}$$

The first integral in the right-hand side of eq. (12.3) can be transformed into a surface integral. Namely, according to formula (I. 44) we have

$$-\mathbf{E} \cdot (\nabla \times \mathbf{H}) - \mathbf{H} \cdot (\nabla \times \mathbf{E}) = \nabla \cdot (\mathbf{E} \times \mathbf{H}).$$

Hence

$$\begin{aligned} \int [\mathbf{E} \cdot (\nabla \times \mathbf{H}) - \mathbf{H} \cdot (\nabla \times \mathbf{E})] dV &= - \int \nabla \cdot (\mathbf{E} \times \mathbf{H}) dV = \\ &= - \oint (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{S}, \end{aligned}$$

and instead of (12.3) one can write

$$\frac{dW}{dt} = \int \mathbf{j} \cdot \mathbf{E} dV = - \frac{c}{4\pi} \oint (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{S} - \frac{\partial}{\partial t} \int \frac{E^2 + H^2}{8\pi} dV. \quad (12.4)$$

Consider the case when the integration volume V increases indefinitely until it covers all space. If the field vectors \mathbf{E} and \mathbf{H} tend to zero more rapidly than $1/r$ as $r \rightarrow \infty$, then the surface integral reduces to zero. Indeed, the integrand decreases more rapidly than $1/r^2$, while the magnitude of the surface increases as r^2 . Then (12.4) reduces to the equality

$$\frac{dW}{dt} = - \frac{d}{dt} \int u_0 dV, \quad (12.5)$$

where

$$u_0 = \frac{E^2 + H^2}{8\pi}. \quad (12.6)$$

Since the left-hand side of (12.5) represents the work done per second, the right-hand side represents the decrease in field energy per unit time.

In a closed system consisting of a field and particles the work done by the electromagnetic field on the particles is equal to the decrease in the energy of the field itself. In this case one has to ascribe to the electromagnetic field an energy whose density u_0 is expressed by formula (12.6). The expression

$$u = \int \frac{E^2 + H^2}{8\pi} dV$$

cannot be reduced to quantities which are determined only by the relative position and motion of charges. Hence it cannot be assumed to be the potential energy of the system of interacting particles. In particular the field energy density differs from zero in a region of space which is free of particles.

The possession of an energy by the electromagnetic field obviously shows that the field can in no way be considered as a mathematical fiction, a convenient method of calculating the interaction between charged particles. On the contrary, the field is as real as the particles. We shall convince ourselves more than once of the reality of the electromagnetic field on the basis of other facts also. However, within the framework of classical electrodynamics the interrelation between the field and the charges remains unexplained. Only quantum electrodynamics, which will be expounded briefly in Part V of this book, allows one to comprehend more profoundly the essence of the interrelation between the field and particles.

Let us now consider a region of the field which has a finite volume V and a limited surface S . Then eq. (12.4), expressing the energy conservation law, shows that the decrease in the field energy per unit time

$$\frac{\partial}{\partial t} \int \frac{E^2 + H^2}{8\pi} dV$$

in a certain volume V is equal to the work done by the field forces dW/dt per unit time on the charges contained in the same volume, together with the flux

$$\frac{c}{4\pi} \oint (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{S},$$

flowing through the closed surface S surrounding the volume V . It is obvious that this flux must be interpreted as an energy flux flowing out of the volume V . The energy flux is the electromagnetic field energy flux, since it also differs from zero when no particles pass through the surface carrying away energy. The electromagnetic field energy flux is characterized by a vector σ , called the Poynting vector, which is equal to

$$\sigma = \frac{c}{4\pi} \mathbf{E} \times \mathbf{H}. \quad (12.7)$$

The Poynting vector σ represents the field energy flux flowing through 1 cm^2 in the direction perpendicular to the field vectors \mathbf{E} and \mathbf{H} , and forms with them a right-handed screw system of coordinates.

Later we shall present a number of examples of the calculation of the Poynting vector. Here we shall confine ourselves only to the following remark. The vector σ is determined formally only with an accuracy to within

the curl of a certain vector \mathbf{a} . Assuming $\boldsymbol{\sigma}' = \boldsymbol{\sigma} + \nabla \times \mathbf{a}$, we have

$$\oint \boldsymbol{\sigma}' \cdot d\mathbf{S} = \oint \boldsymbol{\sigma} \cdot d\mathbf{S} + \oint (\nabla \times \mathbf{a}) \cdot d\mathbf{S} = \oint \boldsymbol{\sigma} \cdot d\mathbf{S},$$

since the integral of $\nabla \times \mathbf{a}$ over a closed surface is always equal to zero.

In reality, however, as will be shown in Part II, $\boldsymbol{\sigma}$ should be interpreted as the density of the field energy flux, assuming $\nabla \times \mathbf{a} = 0$. Sometimes as an example of the inadequacy of such an interpretation of $\boldsymbol{\sigma}$ the case of crossed static electric and magnetic fields is quoted. Formally, in this case $\boldsymbol{\sigma} \neq 0$, although there is no energy flux. However, it is forgotten that the vector $\boldsymbol{\sigma}$ must enter into the energy conservation law expressed by formula (12.4). The latter loses any sense in static fields, if it is assumed that $\boldsymbol{\sigma} \neq 0$.

In deriving formula (12.5) we have assumed that the integral $\oint \boldsymbol{\sigma} \cdot d\mathbf{S}$ reduces to zero in integrating over a closed surface of infinitely large radius. We shall see that in the problems of radiation theory fields are encountered which decrease with increasing distance according to the law $|\mathbf{E}| \sim |\mathbf{H}| \sim 1/r$ as $r \rightarrow \infty$. In this case the integral $\oint \boldsymbol{\sigma} \cdot d\mathbf{S}$, taken over an infinitely distant surface, will have a finite value. Physically this means that the system losing a part of its energy emits radiation.

Writing the energy conservation law in the differential form

$$\frac{\partial}{\partial t} \frac{E^2 + H^2}{8\pi} = -\mathbf{j} \cdot \mathbf{E} - \nabla \cdot \boldsymbol{\sigma}, \quad (12.8)$$

we draw attention to its analogy with the continuity eq. (8.5), expressing the charge conservation law. The left-hand side of formula (12.8) represents the change in the field energy of unit volume (change of a conserved quantity), while on the right-hand side there stands the work produced on the charges contained in this volume, as well as the divergence of the flux of the conserved quantity (the energy density).

§13. Momentum conservation law of the electromagnetic field

The electromagnetic field possesses momentum density as well as energy density.

We shall consider the change in the momentum of the particles confined in a volume V . The calculations are carried out in the same way as in deriving the energy conservation law.

From eq. (8.8) we have

$$\begin{aligned}
\frac{d\mathbf{P}_{\text{part}}}{dt} &= \frac{d}{dt} \int \mathbf{p}_0 dV = \int \rho \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right) dV = \\
&= \int \rho \mathbf{E} dV + \int \frac{1}{c} \mathbf{j} \times \mathbf{H} dV,
\end{aligned} \tag{13.1}$$

where \mathbf{P}_{part} is the total momentum of the particles. Expressing ρ and \mathbf{j} in terms of the field strengths according to (8.4) and (8.3), we find

$$\begin{aligned}
\frac{d}{dt} \mathbf{P}_{\text{part}} &= \frac{1}{4\pi} \int \mathbf{E} (\nabla \cdot \mathbf{E}) dV - \frac{1}{4\pi c} \int \frac{\partial \mathbf{E}}{\partial t} \times \mathbf{H} dV \\
&\quad + \frac{1}{4\pi} \int (\nabla \times \mathbf{H}) \times \mathbf{H} dV.
\end{aligned} \tag{13.2}$$

We symmetrize the above equation by adding to its right-hand side the expression

$$\frac{1}{4\pi} \left(\nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{H}}{\partial t} \right) \times \mathbf{E} + \frac{1}{4\pi} \mathbf{H} (\nabla \cdot \mathbf{H}), \tag{13.3}$$

which is equal to zero. Then we have

$$\begin{aligned}
\frac{d}{dt} \mathbf{P}_{\text{part}} &= - \frac{1}{4\pi c} \frac{\partial}{\partial t} \int \mathbf{E} \times \mathbf{H} dV \\
&\quad + \frac{1}{4\pi} \int \{ \mathbf{E} (\nabla \cdot \mathbf{E}) + \mathbf{H} (\nabla \cdot \mathbf{H}) - \mathbf{E} \times (\nabla \times \mathbf{E}) - \mathbf{H} \times (\nabla \times \mathbf{H}) \} dV.
\end{aligned} \tag{13.4}$$

The second integral can be transformed into a surface integral. This transformation will be carried out below. It is clear that a surface integral containing field vectors to the second power will tend to zero as the surface increases indefinitely, provided the field vectors decrease more rapidly than the function $1/r$. Then, passing over to an infinitely large volume and discarding the second integral in (13.4), we arrive at the expression

$$\mathbf{P}_{\text{part}} + \frac{1}{4\pi c} \int \mathbf{E} \times \mathbf{H} dV = \text{const}. \tag{13.5}$$

Formula (13.5) shows that the total momentum of a closed system con-

sisting of a field and particles is conserved. The quantity

$$\mathbf{g} = \frac{1}{4\pi c} \mathbf{E} \times \mathbf{H} \quad (13.6)$$

represents the momentum density (the momentum per unit volume) of the electromagnetic field. For the interaction between a field and particles there holds, in addition to the total energy conservation law, the total momentum conservation law. The transfer of a momentum to the particles is accompanied by a decrease in the momentum of the field. A momentum loss of the particles (for example, in the emission of radiation) leads to an increase in the momentum of the field.

Now we shall show that the second integral in the form of (13.4) can be reduced to a surface integral. Since the integral

$$\int \{ \mathbf{E}(\nabla \cdot \mathbf{E}) + \mathbf{H}(\nabla \cdot \mathbf{H}) - \mathbf{E} \times (\nabla \times \mathbf{E}) - \mathbf{H} \times (\nabla \times \mathbf{H}) \} dV$$

is symmetric with respect to the vectors \mathbf{E} and \mathbf{H} , we shall consider only the integral

$$I = \int \{ \mathbf{E}(\nabla \cdot \mathbf{E}) - \mathbf{E} \times (\nabla \times \mathbf{E}) \} dV. \quad (13.7)$$

Making use of the vector identities (I. 53) and (I. 48) we can write

$$\int (\mathbf{E} \cdot \nabla) \mathbf{E} dV = \oint (\mathbf{E} \cdot \mathbf{n}) \mathbf{E} dS - \int \mathbf{E}(\nabla \cdot \mathbf{E}) dV, \quad (13.8)$$

$$\int (\mathbf{E} \cdot \nabla) \mathbf{E} \cdot dV = \int \nabla \cdot \frac{1}{2} E^2 dV - \int \mathbf{E} \times (\nabla \times \mathbf{E}) dV. \quad (13.9)$$

Subtracting (13.9) from (13.8) we find

$$\int \{ \mathbf{E}(\nabla \cdot \mathbf{E}) - \mathbf{E} \times (\nabla \times \mathbf{E}) \} dV = - \int \nabla \cdot \frac{1}{2} E^2 dV + \oint (\mathbf{E} \cdot \mathbf{n}) \mathbf{E} dS.$$

Taking into account (I.23), we obtain

$$I = \oint \{ (\mathbf{E} \cdot \mathbf{n}) \mathbf{E} - \mathbf{n} \cdot \frac{1}{2} E^2 \} dS.$$

An analogous expression can be written for the magnetic part of the integral which interests us. Thus, finally,

$$\int \{ \mathbf{E}(\nabla \cdot \mathbf{E}) - \mathbf{E} \times (\nabla \times \mathbf{E}) + \mathbf{H}(\nabla \cdot \mathbf{H}) - \mathbf{H} \times (\nabla \times \mathbf{H}) \} dV = \\ = \oint \{ \mathbf{E} \cdot \mathbf{n} \} \mathbf{E} - \mathbf{n} \frac{1}{2} E^2 + \{ \mathbf{H} \cdot \mathbf{n} \} \mathbf{H} - \mathbf{n} \frac{1}{2} H^2 \} dS. \quad (13.10)$$

Letting the radius of the integration surface go to infinity and assuming that the fields \mathbf{E} and \mathbf{H} decrease at infinity more rapidly than $1/r$, we arrive at the statement expressed earlier that the whole surface integral is equal to zero.

We shall not dwell on the consideration of a more complex case when the integration in (13.4) is carried out over a finite volume *.

We shall only state the result of such a case: the change in the total momentum of a field in a certain volume, $\int \mathbf{g} dV$, is equal to a change in the momentum of the particles confined in the volume and to the momentum flux through the surface bounding the volume.

The theoretical prediction of the existence of field momentum was first confirmed experimentally by P. N. Lebedev in 1901 in the form of the pressure of light. The electromagnetic field momentum is small under ordinary conditions and often lies below the limit of experimental errors. However, in the realm of atomic phenomena the electromagnetic field momentum becomes comparable with the momentum of particles and plays a paramount role in all processes of interaction between radiation and matter. Besides, in the realm of atomic phenomena, the radiation pressure plays a most essential role in processes taking place inside stars and star atmospheres and in other astrophysical phenomena.

It is interesting to note that between the momentum density vector \mathbf{g} and the Poynting vector there is the relation

$$\mathbf{g} = \frac{1}{c^2} \boldsymbol{\sigma}. \quad (13.11)$$

In the chapter devoted to the theory of relativity we shall see that there is a very general relation between the energy and momentum, from which the formula (13.9) is obtained as a consequence.

* See R. Becker, *Electromagnetic fields and interactions* (Blackie, London, 1964); I. E. Tamm, *Osnovy teorii elektrichestva (Introduction to the theory of electricity)* (Gostekhizdat, 1954) p. 506; and in particular, Y. I. Frenkel, *Elektrodinamika (Electrodynamics)* (GTTI, 1934) p. 235, where a more complete interpretation of the expression (13.6) is given.

In addition to the vector of the field momentum density \mathbf{g} one can consider the angular momentum density

$$\mathbf{k}_0 = \mathbf{r} \times \mathbf{g} = \frac{1}{4\pi c} \mathbf{r} \times (\mathbf{E} \times \mathbf{H}).$$

The field angular momentum in a volume V is equal to

$$\mathbf{k} = \frac{1}{4\pi c} \int \mathbf{r} \times (\mathbf{E} \times \mathbf{H}) dV.$$

It can be shown that a conservation law holds for the angular momentum, as well as for the energy and linear momentum. The angular momentum of the electromagnetic field plays an important role in atomic processes. In phenomena on the macroscopic scale the angular momentum has been measured relatively recently *.

* I. E. Tamm, *Osnovy teorii elektrichestva (Introduction to the theory of electricity)* (Gostekhizdat, 1954) p. 502.

The Electrostatic Field

§ 14. The electrostatic field

Having formulated the general equations of the electromagnetic field and discussed the basic consequences following from them, we can proceed to the discussion of particular cases of electromagnetic fields. We shall progress successively from the most simple to more complex cases.

The simplest example of the electromagnetic field is the field of charges at rest.

We write Maxwell's equations for this case

Differential form

$$\nabla \cdot \mathbf{E} = 4\pi\rho, \quad (14.1)$$

$$\nabla \times \mathbf{E} = 0, \quad (14.2)$$

$$\nabla \cdot \mathbf{H} = 0, \quad (14.3)$$

$$\nabla \times \mathbf{H} = 0, \quad (14.4)$$

Integral form

$$\oint \mathbf{E} \cdot d\mathbf{S} = 4\pi \sum e_i, \quad (14.1')$$

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0, \quad (14.2')$$

$$\oint \mathbf{H} \cdot d\mathbf{S} = 0, \quad (14.3')$$

$$\oint \mathbf{H} \cdot d\mathbf{l} = 0. \quad (14.4')$$

The system of equations of the electromagnetic field reduces here to systems of independent equations for the electric and magnetic fields. The solution of the equations for the magnetic field, which does not depend on

time, has the trivial form

$$\mathbf{H} = 0.$$

This means that charges at rest are not surrounded by a magnetic field. The electric field associated with charges at rest is, as we have already mentioned, irrotational. Its sources and sinks are the charges.

In practice one most often needs to find the electric field distribution when the distribution of the charge density in space, $\rho(\mathbf{r})$, is known. For this it is necessary to integrate the system of differential eqs. (14.1) and (14.2) for a given function $\rho(\mathbf{r})$. This is the so-called direct problem of electrostatics.

An incomparably simpler, but more seldom encountered problem of electrostatics is the inverse one — finding the charge density $\rho(\mathbf{r})$ from a given field distribution $\mathbf{E}(\mathbf{r})$. In order to solve the inverse problem of electrostatics, it is sufficient, according to (14.1), to find the divergence of the given field.

As we have already mentioned in §4, to find the general solution of the equations of the electrostatic field it is convenient to make use of the method of electrostatic potential. According to (4.11) or (10.2), it can be assumed that

$$\mathbf{E} = - \nabla \varphi. \quad (14.5)$$

From (14.1) we obtain Poisson's equation:

$$\nabla^2 \varphi = - 4\pi \rho. \quad (14.6)$$

Eq. (14.2) will be satisfied automatically by introducing the potential according to formula (14.5), since for an arbitrary form of the function $\varphi(\mathbf{r})$ there holds the equality

$$\nabla \times \nabla \varphi = 0.$$

Consequently, Maxwell's equations for the electrostatic field (14.1) and (14.2) are completely equivalent to Poisson's equation. Knowing its solution — the scalar potential $\varphi(\mathbf{r})$ — one can determine the strength of the field \mathbf{E} by differentiation.

It should be stressed that only the field strength \mathbf{E} has a physical meaning. The scalar potential is only an auxiliary, though very convenient, quantity. The value of the potential is determined in electrostatics to within an arbitrary constant: adding any constant a to the potential φ , we arrive at a potential

$\varphi' = \varphi + a$, which corresponds to a field $\mathbf{E} = -\nabla \varphi' = -\nabla \varphi$. This transformation is a particular case of the gauge transformation considered in § 11. Because of the incomplete determination of the potential it is senseless to speak about the numerical value of the potential φ at a given point of the field.

Henceforth, in considering solutions of Poisson's equation, and in discussing other properties of the potential, we shall assume a definite behaviour of the potential φ at infinity. If it is assumed that all the charges are distributed in a finite region of space surrounding the point chosen as the origin, then as $r \rightarrow \infty$ the field strength will not decrease more slowly than $1/r^2$. In accordance with this, the solutions of the Poisson equation must satisfy the requirement

$$\varphi \rightarrow 0 \quad \text{as} \quad r \rightarrow \infty. \quad (14.7)$$

From the mathematical standpoint Poisson's equation, a second order partial differential equation, is in some respects more convenient and simpler for the calculation than the field eqs. (14.1) and (14.2), which are first order partial differential equations. If the potential φ at infinity satisfies the condition (14.7), then the solution of Poisson's equation can be written in a general form.

In § 2 the general solution of the eq. (14.6),

$$\begin{aligned} \varphi(\mathbf{r}) &= \int \frac{\rho(\mathbf{r}') dV'}{R} = \int \frac{\rho(\mathbf{r}') dV'}{|\mathbf{r} - \mathbf{r}'|} = \\ &= \int \frac{\rho(x', y', z') dx' dy' dz'}{\sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}}, \end{aligned} \quad (14.8)$$

was given without proof (the latter will be given in § 24). Knowing the distribution of the charge density in space $\rho(x', y', z')$ and integrating over all space, one can find the value of φ at any point (x, y, z) .

The actual calculation of the field according to formula (14.8), requiring the calculation of a triple integral, often turns out to be impracticable. Later, in Part IV, we shall discuss briefly the basic methods of solving the problems of electrostatics taking into account the specific properties of physical bodies — dielectrics and metals. Here we shall confine ourselves only to the simplest system — a system of point charges.

§15. The electrostatic field of a system of point charges

The contrivance of considering the charges as spread out in space and the description of the properties of a system of charges by means of the continuous function $\rho(\mathbf{r}')$ allowed us to pass over from the integral relation (14.1') to the differential equation (14.1). The importance of this transition is clear from the fact that it made it possible to formulate differential field equations. Nevertheless, in some cases it is inadmissible to disregard the point structure of the charges forming real systems. Moreover, in a number of cases it turns out that it is more convenient to carry out the calculations for point systems than for distributed systems.

We write the charge density, characterizing a system of point charges, in the form

$$\rho = \sum e_i \delta(\mathbf{r}' - \mathbf{r}'_i),$$

where \mathbf{r}'_i is the radius-vector of the charge e_i . Substituting this expression into (14.8), we find the field potential of a system of point charges,

$$\varphi = \int \frac{\rho dV'}{|\mathbf{r} - \mathbf{r}'|} = \sum e_i \int \frac{\delta(\mathbf{r}' - \mathbf{r}'_i)}{|\mathbf{r} - \mathbf{r}'|} dV' = \sum \frac{e_i}{R_i}, \quad (15.1)$$

where $R_i = |\mathbf{r} - \mathbf{r}'_i|$, and \mathbf{r} is the radius-vector of the observation point. Here we have made use of the basic property of the delta-function (see III.3). Thus, the solution of the equation

$$\nabla^2 \varphi = -4\pi e \delta(\mathbf{r} - \mathbf{r}_0) \quad (15.2)$$

is the function

$$\varphi = \frac{e}{|\mathbf{r} - \mathbf{r}_0|}. \quad (15.3)$$

Formula (15.3) represents a useful relation which we shall use in what follows.

The field of a system of point charges is given by the formula

$$\mathbf{E} = -\nabla \varphi = \sum \frac{e_i}{R_i^3} \mathbf{R}_i. \quad (15.4)$$

In the case of a single charge, formula (15.4) gives

$$\mathbf{E} = \frac{e}{R^3} \mathbf{R}. \quad (15.5)$$

For the force acting on a test charge ϵ placed in the field of a single charge one obtains the Coulomb law

$$\mathbf{F} = \frac{\epsilon e}{R^3} \mathbf{R}. \quad (15.6)$$

If the number of charges in the system is large, the sums in formula (15.1) and (15.4) contain a large number of terms, and these formulae become of little use for practical calculations. However, formula (15.1) allows a substantial simplification at distances from the system which substantially exceed its spatial extension. Distances which are large in comparison with the dimensions of the system will be called, for brevity, large distances. If the observation point N is located at a large distance from the system, then there holds the inequality (fig. 1.6)

$$|\mathbf{r}| \gg |\mathbf{r}'_i|.$$

Let us consider one of the terms in formula (15.1). In order not to encumber subsequent formulae by indices, we shall not write the sign of the sum, and the distance from the i th charge to the observation point will be written

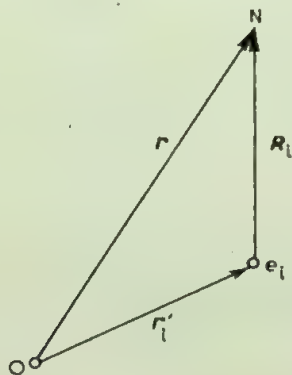


Fig. 1.6

in the form

$$\begin{aligned} \frac{1}{R_i} &= \frac{1}{|\mathbf{r} - \mathbf{r}'_i|} = \frac{1}{\sqrt{(x - x'_i)^2 + (y - y'_i)^2 + (z - z'_i)^2}} = \\ &= \frac{1}{\sqrt{\sum_{\alpha=1}^3 (x_{\alpha} - x'_{i\alpha})^2}}, \end{aligned} \quad (15.7)$$

where the index α signifies the three components of the corresponding vectors. Since $|x'_{i\alpha}| \ll |x_{\alpha}|$, expanding (15.7) in the Taylor series we have

$$\begin{aligned} \frac{1}{R_i} &= \frac{1}{|\mathbf{r} - \mathbf{r}'_i|} = \frac{1}{r} - \sum_{\alpha} x'_{i\alpha} \left\{ \frac{\partial}{\partial x_{\alpha}} \left(\frac{1}{r} \right) \right\}_0 + \\ &+ \frac{1}{2} \sum_{\alpha, \beta} x'_{i\alpha} x'_{i\beta} \left\{ \frac{\partial^2}{\partial x_{\alpha} \partial x_{\beta}} \left(\frac{1}{r} \right) \right\}_0 + \dots \end{aligned} \quad (15.8)$$

Substituting the expansion (15.8) into formula (15.1), we find

$$\begin{aligned} \varphi &= \frac{\sum e_i}{r} - \sum_{\alpha} \left(\sum_i e_i x'_{i\alpha} \right) \frac{\partial}{\partial x_{\alpha}} \left(\frac{1}{r} \right) + \\ &+ \frac{1}{2} \sum_{\alpha, \beta} \sum_i (e_i x'_{i\alpha} x'_{i\beta}) \frac{\partial^2}{\partial x_{\alpha} \partial x_{\beta}} \left(\frac{1}{r} \right) + \dots = \\ &= \varphi_0 + \varphi_1 + \varphi_2 + \dots, \end{aligned} \quad (15.9)$$

where

$$\varphi_0 = \frac{\sum e_i}{r} = \frac{e}{r}, \quad (15.10)$$

$$\varphi_1 = - \sum_{\alpha} \left(\sum_i e_i x'_{i\alpha} \right) \frac{\partial}{\partial x_{\alpha}} \left(\frac{1}{r} \right), \quad (15.11)$$

$$\varphi_2 = \frac{1}{2} \sum_{\alpha, \beta} \left(\sum_i e_i x'_{i\alpha} x'_{i\beta} \right) \frac{\partial^2}{\partial x_{\alpha} \partial x_{\beta}} \left(\frac{1}{r} \right). \quad (15.12)$$

The summation with respect to i is carried out over all charges of the system. Hence $\sum_i e_i = e$ represents the total charge of the system.

We see that at large distances the ratio between two successive terms of the expansion of the potential is of the order of magnitude of the ratio $\frac{\text{size of the system}}{\text{distance to the observation point}}$. The first term of the expansion φ_0 is the same as the potential of the field produced at the given point N by a charge e equal to the total charge of the system. Every subsequent term of the expansion contains the above mentioned ratio, of size of the system ($\sim |\mathbf{x}'|$) to the distance to the observation point $|\mathbf{r}|$, to a higher power.

In the case of an electrically neutral system the total charge $e = \sum_i e_i = 0$, and the first term of the series (15.9) vanishes. We shall deal with such systems very often. It is sufficient to point out, for example, that all atoms and molecules are electrically neutral systems. The potential of the field produced by an electrically neutral system of charges is given by the expansion (15.9) beginning with the second term φ_1 . Let us consider it in more detail. We write φ_1 in the vector form

$$\begin{aligned}\varphi_1 &= -\sum_{\alpha} \left(\sum_i e_i x'_{i\alpha} \right) \frac{\partial}{\partial x_{\alpha}} \left(\frac{1}{r} \right) = -\sum_i e_i \left(\nabla \frac{1}{r} \right) \cdot \mathbf{r}'_i = \\ &= -\left(\nabla \frac{1}{r} \right) \cdot \sum_i e_i \mathbf{r}'_i,\end{aligned}\quad (15.13)$$

where the gradient is taken with respect to the coordinates of the observation point. The quantity

$$\mathbf{d} = \sum e_i \mathbf{r}'_i = \int \rho \mathbf{r}' dV' \quad (15.14)$$

is called the dipole moment of the system. In the particular case of a system consisting of two charges of the same magnitude but opposite sign, which is called a dipole, the dipole moment is equal to

$$\mathbf{d} = e_1 \mathbf{r}'_1 + e_2 \mathbf{r}'_2 = |e| (\mathbf{r}'_1 - \mathbf{r}'_2),$$

i.e. to the product of the value of the charge and the vector $\mathbf{l} = (\mathbf{r}'_1 - \mathbf{r}'_2)$.

The field of an electrically neutral system in the first approximation (called the dipole approximation) is written in the form

$$\varphi \approx \varphi_1 = -\mathbf{d} \cdot \left(\nabla \frac{1}{r} \right) = \frac{\mathbf{d} \cdot \mathbf{r}}{r^3} = \frac{d \cos \theta}{r^2}, \quad (15.15)$$

where θ is the angle between the dipole moment and the radius-vector drawn to the observation point. Thus, the potential of the field of an electrically neutral system decreases (at large distances from the system) according to the law $\varphi \sim 1/r^2$.

Just as we have passed over from point charges to a charge distributed continuously in space with density ρ , one can introduce the notion of the dipole moment density \mathbf{p} distributed continuously in space. By definition, \mathbf{p} is the dipole moment per unit volume. The potential of the field produced by the overall system can, obviously, be written in the form

$$\varphi = -\int \mathbf{p} \cdot \left(\nabla \frac{1}{r} \right) dV',$$

where $\mathbf{p} \cdot (\nabla(1/r))dV'$ is the potential produced at the point N by the dipole moment contained in a volume dV' , and the integration is carried out over the entire volume of the system. It should be noted that instead of the differentiation with respect to the coordinates of the observation point one often makes use of the differentiation with respect to the coordinates of the source. Then according to (1.17) we have $\nabla(1/r) = -\nabla'(1/r)$ and instead of (15.15) we can write

$$\varphi = \mathbf{d} \cdot \left(\nabla' \frac{1}{r} \right), \quad (15.16)$$

or

$$\varphi = \int \mathbf{p} \cdot \left(\nabla' \frac{1}{r} \right) dV. \quad (15.17)$$

We shall need formula (15.17) later (see Part IV).

Consider the problem of the dependence of the dipole moment on the choice of the origin. Assume that we locate a new origin at an arbitrary constant vector \mathbf{a} with respect to the old origin, i.e. we perform the transformation

$$\mathbf{r}'_i = \mathbf{r}''_i + \mathbf{a}.$$

Then the dipole moment will be equal to

$$\mathbf{d} = \sum e_i \mathbf{r}'_i = \sum e_i \mathbf{r}''_i + \sum e_i \mathbf{a} = \mathbf{d}' + \mathbf{a} \sum e_i,$$

where $\mathbf{d}' = \sum e_i \mathbf{r}''_i$. If the system as a whole is electrically neutral, then $\sum e_i = 0$ and $\mathbf{d}' = \mathbf{d}$. In this case the value of the dipole moment does not change when the origin is displaced. If, on the contrary, the system possesses a total charge, then $\mathbf{d}' \neq \mathbf{d}$ and, consequently, the dipole moment of the system depends on the choice of the origin. It is then always possible to find a value of \mathbf{a} such that the dipole moment reduces to zero. Thus, the dipole moment of any system possessing a total charge may be considered as equal to zero.

We now determine the field of an electrically neutral system in the dipole approximation.

We have

$$\begin{aligned} \mathbf{E} &= -\nabla\varphi \approx -\nabla\varphi_1 = -\nabla\left(\frac{\mathbf{d} \cdot \mathbf{r}}{r^3}\right) = \\ &= -\frac{1}{r^3} \nabla(\mathbf{d} \cdot \mathbf{r}) - (\mathbf{d} \cdot \mathbf{r}) \left(\nabla \frac{1}{r^3}\right). \end{aligned}$$

Calculation by means of formulae (1.47) gives

$$\mathbf{E} = \frac{3\mathbf{r}(\mathbf{r} \cdot \mathbf{d}) - r^2 \mathbf{d}}{r^5}. \quad (15.18)$$

At large distances the field of an electrically neutral system decreases according to the law $\mathbf{E} \sim 1/r^3$ and has a strongly pronounced asymmetry. In polar coordinates (r, θ) its components (according to 1.71) have the form

$$\begin{aligned} E_r &= -\frac{\partial\varphi}{\partial r} = \frac{2d \cos \theta}{r^3} = \text{radial component}, \\ E_\theta &= -\frac{1}{r} \frac{\partial\varphi}{\partial \theta} = \frac{d \sin \theta}{r^3} = \text{meridional component}. \end{aligned}$$

§ 16. Quadrupole moment

If the dipole moment of an electrically neutral system of charges is equal

to zero, then in the expansion of the potential (15.9) one has to take into account the expansion term φ_2 .

An example of an electrically neutral system with dipole moment zero is the system of two dipoles of the same magnitude with opposite directions of the dipole moments placed an infinitely small distance apart. Such a system is called a quadrupole.

The potential of the field produced by a quadrupole has the form

$$\varphi \approx \varphi_2 = \frac{1}{2} \sum_{\alpha, \beta} \sum_i e_i x'_{i\alpha} x'_{i\beta} \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \left(\frac{1}{r} \right). \quad (16.1)$$

In order to obtain φ one has to calculate the expression

$$\begin{aligned} \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \left(\frac{1}{r} \right) &= \frac{\partial}{\partial x_\alpha} \left(\frac{\partial}{\partial x_\beta} \frac{1}{r} \right) = - \frac{\partial}{\partial x_\alpha} \frac{x_\beta}{r^3} = \\ &= -x_\beta \frac{\partial}{\partial x_\alpha} \left(\frac{1}{r^3} \right) - \frac{1}{r^3} \frac{\partial x_\beta}{\partial x_\alpha} = -\frac{1}{r^3} \delta_{\alpha\beta} + \frac{3x_\alpha x_\beta}{r^5}, \end{aligned}$$

where $\delta_{\alpha\beta}$ is the Kronecker symbol, α, β assume the values 1, 2, 3, and x_α, x_β represent an abbreviated notation of the coordinates $x_1 = x, x_2 = y, x_3 = z$. We then have

$$\varphi_2 = \frac{1}{2} \sum_{\alpha, \beta} \sum_i e_i x'_{i\alpha} x'_{i\beta} \left(\frac{3x_\alpha x_\beta}{r^5} - \frac{\delta_{\alpha\beta}}{r^3} \right).$$

The whole set of the quantities $\sum_i e_i x'_{i\alpha} x'_{i\beta}$ is a second-rank tensor. This tensor is called the quadrupole moment of the system and is denoted by $D_{\alpha\beta}$. Then

$$\varphi_2 = \frac{1}{2} \sum_{\alpha, \beta} D_{\alpha\beta} \left(\frac{3x_\alpha x_\beta}{r^5} - \frac{\delta_{\alpha\beta}}{r^3} \right), \quad (16.2)$$

$$D_{\alpha\beta} = \sum_i e_i x'_{i\alpha} x'_{i\beta}. \quad (16.3)$$

If one passes over to a continuous distribution of charge, the quadrupole moment can be written in the form

$$D_{\alpha\beta} = \int \rho x'_\alpha x'_\beta dV'. \quad (16.3')$$

Omitting for brevity the index of summation over all the particles, we can write the expression for φ_2 in the coordinate representation:

$$\varphi_2 = \frac{1}{2} \sum e \left\{ x'^2 \left(\frac{3x^2}{r^5} - \frac{1}{r^3} \right) + y'^2 \left(\frac{3y^2}{r^5} - \frac{1}{r^3} \right) + z'^2 \left(\frac{3z^2}{r^5} - \frac{1}{r^3} \right) + \right. \\ \left. + 2x'y' \frac{3xy}{r^5} + 2x'z' \frac{3xz}{r^5} + 2y'z' \frac{3yz}{r^5} \right\}.$$

This expression is usually rearranged by adding to it the quantity

$$\frac{1}{3} r'^2 \frac{3}{r^3} \left(\frac{x^2 + y^2 + z^2}{r^2} - 1 \right),$$

which is equal to zero, and by re-grouping the terms in the form

$$\begin{aligned} \varphi_2 &= \frac{1}{2} \sum e \left\{ \left(x'^2 - \frac{1}{3} r'^2 \right) \left(\frac{3x^2}{r^5} - \frac{1}{r^3} \right) + \right. \\ &\quad + \left(y'^2 - \frac{1}{3} r'^2 \right) \left(\frac{3y^2}{r^5} - \frac{1}{r^3} \right) + \left(z'^2 - \frac{1}{3} r'^2 \right) \left(\frac{3z^2}{r^5} - \frac{1}{r^3} \right) + \\ &\quad + 2x'y' \frac{3xy}{r^5} + 2x'z' \frac{3xz}{r^5} + 2y'z' \frac{3yz}{r^5} \Big\} = \\ &= \frac{1}{2} \sum_{\alpha, \beta} D_{\alpha\beta} \left(\frac{3x_{\alpha}x_{\beta}}{r^5} - \frac{\delta_{\alpha\beta}}{r^3} \right). \end{aligned} \quad (16.4)$$

In this case the quadrupole moment $D_{\alpha\beta}$ is defined as

$$D_{\alpha\beta} = \sum_i e_i \left(x'_{i\alpha} x'_{i\beta} - \frac{1}{3} r'^2 \delta_{\alpha\beta} \right). \quad (16.5)$$

The set of the quantities $D_{\alpha\beta}$ is easily written in the explicit form:

$$\begin{aligned} D_{xx} &= \sum e_i [(x'_i)^2 - \frac{1}{3} r'^2_i], & D_{xy} &= D_{yx} = \sum e_i x'_i y'_i, \\ D_{yy} &= \sum e_i [(y'_i)^2 - \frac{1}{3} r'^2_i], & D_{yz} &= D_{zy} = \sum e_i y'_i z'_i, \\ D_{zz} &= \sum e_i [(z'_i)^2 - \frac{1}{3} r'^2_i], & D_{zx} &= D_{xz} = \sum e_i z'_i x'_i. \end{aligned}$$

All of the nine quantities $D_{\alpha\beta}$ forming the quadrupole moment depend, obviously, only on the positions and values of the charges of the system.

From the definition (16.5) it is clear that the quadrupole moment tensor is symmetric, so that

$$D_{\alpha\beta} = D_{\beta\alpha}.$$

The symmetric second-rank tensor has six independent components.

Further we note that the sum of all diagonal components of the quadrupole moment is equal to zero:

$$D_{xx} + D_{yy} + D_{zz} = 0. \quad (16.6)$$

This reduces the number of independent components of the quadrupole moment to five.

Like any symmetric tensor, $D_{\alpha\beta}$ can be reduced to principal axes. This procedure is completely analogous to the reduction of the inertia tensor to principal axes in mechanics. Namely, we perform a rotation of the coordinate frame. Then to the coordinates x'_i, y'_i, z'_i there will correspond new coordinates x_i, y_i, z_i . We choose the coefficients of the corresponding linear transformation in such a way that the components of $D_{\alpha\beta}$ with different values of the indices α and β reduce to zero. It can be shown that such a choice of the coefficients is always possible.

With the new axes we shall have:

$$D_1 = \frac{1}{3} \sum e_i (2x_{i1}^2 - x_{i2}^2 - x_{i3}^2) = D_{11},$$

$$D_2 = \frac{1}{3} \sum e_i (2x_{i2}^2 - x_{i1}^2 - x_{i3}^2) = D_{22},$$

$$D_3 = -(D_1 + D_2) = D_{33}.$$

An important case is the system of charges whose arrangement possesses axial symmetry. Let the symmetry axis be the x_3 -axis. The condition of symmetry with respect to the x_3 -axis allows one to write

$$\sum e_i x_{i1}^2 = \sum e_i x_{i2}^2,$$

so that the quadrupole moment has the components

$$D_1 = \frac{1}{3} \sum e_i (x_{i1}^2 - x_{i3}^2) = -\frac{1}{2} D,$$

$$D_2 = \frac{1}{3} \sum e_i (x_{i2}^2 - x_{i3}^2) = -\frac{1}{2} D,$$

$$D_3 = D.$$

The sign of the quantity D is called the sign of the quadrupole moment. According to (16.4), the potential of the quadrupole field is equal to

$$\varphi_2 = -\frac{3}{4} D \frac{r^2 - 3x_3^2}{r^5} = -\frac{3}{4} D \frac{1 - 3 \cos^2 \theta}{r^3},$$

where θ is the angle between the symmetry axis x_3 and the radius-vector \mathbf{r} of the observation point. The general character of the law of decrease of the quadrupole potential with increasing distance r is

$$\varphi_2 \sim 1/r^3.$$

Correspondingly, the electric field decreases according to the law

$$|\mathbf{E}| \sim 1/r^4.$$

As we have already mentioned in the preceding paragraph, if a system of charges as a whole is not electrically neutral, then the value of the dipole moment depends on the choice of the origin. The same also holds for the quadrupole moment. For the case of a system with a total charge differing from zero it is convenient to locate the origin at a point with a coordinate

$$\mathbf{r}_0 = \frac{\sum e_i \mathbf{r}_i}{\sum e_i}.$$

This point can be called the centre of charge of the system. If the origin is located at the centre of charge, then the dipole moment of the system of charges reduces automatically to zero. However, this does not hold for its quadrupole moment. Namely, if the disposition of the charges in the system is not spherically symmetric, then some or all components of the quadrupole moment differ from zero. Hence the presence of a quadrupole moment in a system of charges allows one to infer something about the symmetry of the system. Thus, for example, the presence of axial symmetry leads to the field distribution written above.

In connection with this fact the discovery of a quadrupole-moment in a number of atomic nuclei is of great significance. The presence of the quadrupole moment of nuclei showed that their form is non-spherical.

If the distribution of charges in a system possesses a very high symmetry, its quadrupole moment may turn out to be equal to zero. As an example, we cite a system of eight charges situated at the vertices of an infinitely small parallelepiped with a regular alternation of the signs of the charges. Such a system of charges, called an octupole, has neither dipole nor quadrupole moment. The potential of the octupole field is obtained by taking into account the fourth term of the expansion (15.9).

Continuing the process of expansion (15.9) one can obtain the field potential of multipoles of an arbitrary order.

§17. Work and energy in an external electrostatic field

According to the above, the work of displacement of a test charge from one point of a field to another can be expressed in terms of the change in the potential in the form

$$\begin{aligned} W &= \int_1^2 \mathbf{F} \cdot d\mathbf{l} = \epsilon \int_1^2 \mathbf{E} \cdot d\mathbf{l} = -\epsilon \int_1^2 \nabla \varphi \cdot d\mathbf{l} = \\ &= \epsilon [\varphi(\mathbf{r}_1) - \varphi(\mathbf{r}_2)] = -\epsilon \Delta \varphi. \end{aligned} \quad (17.1)$$

If a displacement of the charges of a system takes place, then the work of the displacement by vectors $d\mathbf{l}_i$ is equal to

$$dW = \sum e_i \mathbf{E} \cdot d\mathbf{l}_i.$$

Later we shall need the expression for the rate at which work is done on a system of charges. We find

$$\frac{dW}{dt} = \sum e_i \mathbf{E} \cdot \frac{d\mathbf{l}_i}{dt} = \sum e_i \mathbf{E} \cdot \mathbf{v}_i. \quad (17.2)$$

For the case of distributed charges

$$\frac{dW}{dt} = \int \rho \mathbf{v} \cdot \mathbf{E} dV = \int \mathbf{j} \cdot \mathbf{E} dV. \quad (17.3)$$

Knowing the work of displacement of a test charge, we can write its potential energy in the electrostatic field in the form

$$-\delta U(\mathbf{r}) = \delta W = -e\delta\varphi(\mathbf{r}),$$

where $U(\mathbf{r})$ is the potential energy at the point \mathbf{r} , and $\varphi(\mathbf{r})$ is the electrostatic potential at the same point. The form of the potential energy does not depend on the choice of the coordinate frame. Hence the relation

$$U = e\varphi \quad (17.4)$$

is valid not only in Cartesian coordinates but also in any generalized coordinates q_i . The generalized forces acting on the test charge can be written in the form

$$Q_i = -e \frac{\partial \varphi}{\partial q_i}. \quad (17.5)$$

Formulae (17.4) and (17.5) can be applied to the case where instead of a test charge an arbitrary system of charges is placed in an external field. In this case it is assumed that the external field \mathbf{E} is strong in comparison with the field produced by the charges of the system. It is assumed moreover that the potential of the external field varies sufficiently slowly from point to point. The potential energy of the system of charges can be written in the form

$$U = \sum e_i \varphi(\mathbf{r}'_i), \quad (17.6)$$

where $\varphi(\mathbf{r}'_i)$ is the potential of the external field at a point \mathbf{r}'_i . Choosing the origin inside the system, we can write the potential in the form

$$\varphi(\mathbf{r}'_i) = \varphi(x', y', z'),$$

where x', y', z' are the distances from the origin to the charge. We now make use of the slowness of the variation of the potential of the external field in the region of space occupied by the charges. The slowly varying function φ can be expanded in a series in the quantities x', y', z' characterizing the extension of the system, and we can confine ourselves to the first terms of the expansion. This gives

$$\begin{aligned}\varphi(\mathbf{r}') &\approx \varphi(0, 0, 0) + x' \frac{\partial \varphi}{\partial x} + y' \frac{\partial \varphi}{\partial y} + z' \frac{\partial \varphi}{\partial z} = \\ &= \varphi(0) + \mathbf{r}' \cdot \nabla \varphi = \varphi(0) - \mathbf{r}' \cdot \mathbf{E}(0).\end{aligned}$$

Here $\varphi(0)$ and $\mathbf{E}(0)$ are respectively the potential and strength of the external field at the origin. Substituting the last expression for $\varphi(\mathbf{r}')$ into (17.6), we find

$$\begin{aligned}U &= \sum e_i \varphi_i \approx \sum e_i \varphi(0) - \sum e_i \mathbf{r}'_i \cdot \mathbf{E}(0) = \\ &= \varphi(0) \sum e_i - \mathbf{E}(0) \cdot \sum e_i \mathbf{r}'_i = e \varphi(0) - \mathbf{E}(0) \cdot \mathbf{d}.\end{aligned}\quad (17.7)$$

In the first approximation the potential energy of a system of charges in an external field is equal to the energy of one charge of the value $e = \sum e_i$, located at the origin. In the case of an electrically neutral system $e = 0$ and

$$U = -\mathbf{d} \cdot \mathbf{E} = -dE \cos \theta, \quad (17.8)$$

where θ is the angle between the dipole moment of the system and the external field vector.

Let us find the generalized forces acting on a system (assuming the latter to be undeformable, so that the distribution of the charges in the system is fixed). The generalized force corresponding to coordinates x, y, z is equal to

$$\mathbf{F} = -\nabla U = \nabla(\mathbf{d} \cdot \mathbf{E}),^*$$

or, evaluating the gradient of the product by means of formula (I.47) and taking into account that \mathbf{d} is a constant vector, we obtain

$$\mathbf{F} = (\mathbf{d} \cdot \nabla) \mathbf{E} + \mathbf{d} \times (\nabla \times \mathbf{E}) = (\mathbf{d} \cdot \nabla) \mathbf{E}. \quad (17.9)$$

In a uniform field ($\mathbf{E} = \text{const}$) an electrically neutral system with a dipole moment is not acted upon by any forces tending to displace it in space. Such forces exist only in a field which is non-uniform.

The generalized force corresponding to the generalized coordinate θ , which determines the orientation of the dipole moment vector, according to a well-known proposition of classical mechanics * represents a couple:

* Landau and E. M. Lifshitz, *Mechanics* (Pergamon, Oxford, 1960).

$$M = - \frac{\partial U}{\partial \theta} = dE \sin \theta . \quad (17.10)$$

The couple tends to turn the system in such a way that its dipole moment will be orientated parallel to the field.

The formulae found above allow one to find easily the law of charge-dipole and dipole-dipole interaction. For this $E(0)$ is understood to be the field produced at the point 0 by the charge or the dipole respectively.

For the potential energy of the charge-dipole interaction we find

$$U = - \mathbf{d} \cdot \mathbf{E} = - \frac{e \mathbf{r} \cdot \mathbf{d}}{r^3}, \quad (17.11)$$

where \mathbf{r} is the vector directed from the charge to the system, the value of which is equal to the distance from the charge to the system (in this approximation we have to disregard the spatial dimensions of the system).

The dipole-dipole potential energy, according to (15.18), is

$$U = - \mathbf{d}_1 \cdot \mathbf{E}_2 = \frac{(\mathbf{d}_1 \cdot \mathbf{d}_2)r^2 - 3(\mathbf{d}_1 \cdot \mathbf{r})(\mathbf{d}_2 \cdot \mathbf{r})}{r^5}, \quad (17.12)$$

where \mathbf{r} is the vector connecting the two dipoles.

§ 18. The interaction energy of a system of charges and the electrostatic field energy

We pass on to the calculation of the energy of a system of interacting charges. This energy can be calculated most simply in the following way. Let a charge e_1 be fixed at a certain point of space. A charge e_2 , which was initially at infinity, is displaced to a point located at a distance r_{12} from the first charge. In this case the external source must do work against the forces of the field:

$$W_{12} = e_2 [\varphi_1(r_{12}) - \varphi_1(r \rightarrow \infty)] = e_2 \varphi_1(r_{12}).$$

Since the potential of the field of the first charge at infinity is equal to zero, $\varphi_1(r_{12})$ represents the potential of the field of the first charge at the point r_{12} , which is equal to

$$\varphi_1(r_{12}) = e_1/r_{12}.$$

Hence the work of displacement of the second charge is equal to

$$W_{12} = \frac{e_1 e_2}{r_{12}}.$$

If a third charge is added to a system of two charges, it is necessary to produce the work

$$W_{123} = \frac{e_1 e_3}{r_{13}} + \frac{e_2 e_3}{r_{23}}.$$

Continuing with such a procedure for a system of N charges, we find that for this one has to expend the work

$$\begin{aligned} W = \frac{1}{2} e_1 \left[\frac{e_2}{r_{12}} + \frac{e_3}{r_{13}} + \dots + \frac{e_N}{r_{1N}} \right] + \frac{1}{2} e_2 \left[\frac{e_1}{r_{21}} + \frac{e_3}{r_{23}} + \dots + \right. \\ \left. + \frac{e_N}{r_{2N}} \right] + \dots = \frac{1}{2} \sum_i \sum_k \frac{e_i e_k}{r_{ik}} \quad (i \neq k). \end{aligned} \quad (18.1)$$

The coefficient $\frac{1}{2}$ is introduced because in the double sum the same terms, corresponding to each pair of particles, for example $e_1 e_2 / r_{12}$ and $e_2 e_1 / r_{21}$ are encountered twice. In order not to introduce a complicated restriction upon the performance of the summation, in (18.1) all such terms are taken into account and the result is reduced by a factor of two.

Introducing φ_i which is the potential produced by all the charges except the i th one at the locus of the latter, we can rewrite (18.1) in the form

$$W = \frac{1}{2} \sum e_i \varphi_i.$$

The work done in forming the charge distribution is equal to the potential energy stored in the system of particles. Thus,

$$U = \frac{1}{2} \sum e_i \varphi_i = \frac{1}{2} \sum e_i e_k / r_{ik}. \quad (18.2)$$

Passing over from point charges to a continuous charge density distribution, one can write (18.2) in the form

$$U = \frac{1}{2} \int \rho \varphi dV = \frac{1}{2} \iint \frac{\rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV dV'. \quad (18.3)$$

The potential energy of the interaction (18.2) is determined by the instantaneous positions of all charges in the system. Formulae (18.2) and (18.3) can be interpreted in the following way: each charge contained in the system possesses a potential energy $\frac{1}{2} e_i \varphi_i$, and the energy of the system is made up of the energies of the charges constituting it.

We now transform formula (18.3), making use of the field equations. Expressing ρ in terms of \mathbf{E} according to (14.1), we find by means of (18.3)

$$\begin{aligned} U &= \frac{1}{8\pi} \int \varphi \nabla \cdot \mathbf{E} dV = \frac{1}{8\pi} \left[\int \nabla \cdot (\varphi \mathbf{E}) dV - \int \mathbf{E} \cdot (\nabla \varphi) dV \right] = \\ &= \int \frac{E^2}{8\pi} dV + \frac{1}{8\pi} \oint \varphi \mathbf{E} \cdot d\mathbf{S}. \end{aligned}$$

The integral over an infinite surface vanishes, since as $r \rightarrow \infty$ we have

$$\varphi < 1/r, \quad E < 1/r^2, \quad S \sim r^2.$$

Hence we find finally

$$U = \int \frac{E^2}{8\pi} dV. \quad (18.4)$$

Formula (18.4) is completely equivalent to (18.3). However, it does not contain any quantities characterizing the electric charges.

It is quite clear that the expression (18.4) is a particular case of the general expression for the electromagnetic field energy, and its derivation is a particular case of the proof presented in §12. However, it is inherent within the framework of electrostatics that neither of the two alternative formulae (18.3) and (18.4) is more fundamental. Since formula (18.3) does not contain any field characteristics, then in electrostatics the field can be treated as an auxiliary, purely mathematical method of describing the interaction between particles. The state of the system and its energy in electrostatics are determined solely by the values of the charges and their relative positions.

We have already emphasized earlier that in the general case of a system of moving charges and fields varying in time the situation differs radically from that in electrostatics. The electromagnetic field cannot in general be treated

as a purely mathematical concept. It is a physical object as real as charged particles.

It is interesting to apply formula (18.4) to a single elementary charge — an electron or proton. Its energy is equal to

$$U = e\varphi(0),$$

where $\varphi(0)$ is the field potential at the point at which the charge is located. Since the field considered is the field of the charge itself, its potential $\varphi = e/r$ increases indefinitely as r tends to zero. This means that a point particle would have an infinitely large self energy.

Thus, the concept of particles as point objects having no spatial extension leads to a physically senseless result. In this connection a number of attempts have been made to construct an electrodynamic theory of elementary particles having finite dimensions (theory of the extended electron). As we shall see in Part II, this theory turned out to be incompatible with the basic propositions of the theory of relativity.

In the problem of the self energy of an elementary charge, classical electrodynamics encountered an insurmountable difficulty. It was clear that the laws of classical electrodynamics, which had been in excellent agreement with experimental facts in the field of macroscopic physics, had a limited region of applicability. In going down to very small distances they must undergo fundamental changes. We shall speak about the limits of applicability of classical electrodynamics in later paragraphs of Part I.

The Quasistationary Magnetic Field

§ 19. The field of a system of charges undergoing a slow quasistationary motion

Next in degree of complexity is the case of the field of charges performing a slow quasistationary motion.

We shall call the motion of a system of charges slow when the velocities $|v|$ are small in comparison with the value of c , the unique characteristic quantity with the dimension of velocity contained in Maxwell's equations. We shall see later that the velocity c represents the velocity of propagation in space of all electromagnetic interactions.

Thus, the assumption of a slow motion of charges means that the finite velocity of propagation of electromagnetic fields can be disregarded (see § 23). For a slow motion it can be assumed that the field at any instant of time is determined approximately by the instantaneous disposition of the charges.

We shall understand quasistationary motion to be a motion of charges in a certain limited region which the charges do not leave during the motion. In this region the charges may move periodically or non-periodically. In the latter case, however, over a very long time the particles will inevitably either pass through the same sequences of states as in a periodic motion, or, in any case, through sequences of similar states. In other words, the motion will be almost periodic. It will be shown below that under these conditions the time

derivatives of fields in Maxwell's equations are small compared with spatial derivatives. Hence the term quasistationary motion.

Since the particles cannot go out of the region, the following condition must be fulfilled at the surface S_1 bounding the region:

$$j_n = 0. \quad (19.1)$$

Here j_n is the component of the current density normal to the surface.

For a slow motion of charges the variation of the charge density in time can be considered small, i.e. it can be assumed that

$$\nabla \cdot \mathbf{j} = 0. \quad (19.2)$$

Thus, for a quasistationary motion the current density vector has a solenoid character. In other words, the quasistationary character of the motion of the charges allows one to represent their trajectories in the form of closed tubes or threads. Each of these tubes is closed on itself inside the region of motion. Such a representation is particularly obvious in the case of macroscopic direct currents flowing, for example, along closed conductors (see § 17, Part IV).

For every closed tube of current there holds the equality

$$\mathbf{j} dV = \mathbf{j} dS dl = j dS dl = dl dl, \quad (19.3)$$

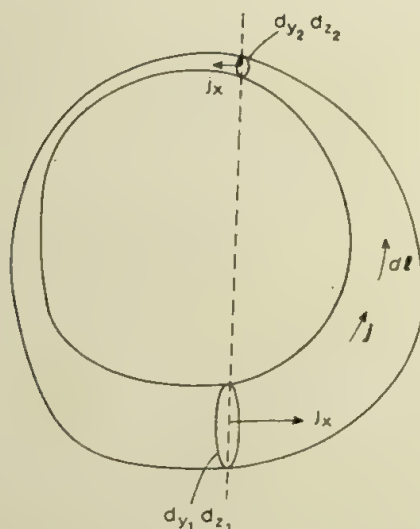


Fig. 1.7

where dS is the cross section, $j dS$ is the direct current flowing through the cross section of the tube, and dl is an element of the tube length. The directions of the vectors \mathbf{j} and $d\mathbf{l}$ are obviously the same. The integral of the current density over the entire volume is

$$\int \mathbf{j} dV = \int \oint dI d\mathbf{l} = \int dI \oint d\mathbf{l} = 0, \quad (19.4)$$

since the integral along the closed tube is $\oint d\mathbf{l} = 0$. The meaning of this equality is very simple. Let us consider, for example, its x -component. The integral $\int j_x dy dz$ represents the total current through the plane (yz) crossing the current tube (fig. 1.7). In a quasistationary state of the system the total current through any cross section is equal to zero. The number of charges passing along the normal to the cross section through all current tubes in both directions must be the same, since the charges are moving in a limited region of space.

We now pass over to the formulation of Maxwell's equations for the field of a system of charges performing a slow quasistationary motion.

In order to find out which simplifications can be introduced into Maxwell's equations for such a motion, we shall estimate the order of magnitude of the terms of the equations. Such methods of estimation are widely used in theoretical physics.

We begin with the estimation of the time derivatives appearing in Maxwell's equations. Since the system considered performs a periodic or almost periodic motion, the order of magnitude of the quantities $|\partial \mathbf{E} / \partial t|$ and $|\partial \mathbf{H} / \partial t|$ can be estimated by writing

$$\left| \frac{\partial \mathbf{E}}{\partial t} \right| \sim \frac{E}{T}, \quad \left| \frac{\partial \mathbf{H}}{\partial t} \right| \sim \frac{H}{T},$$

where T is the characteristic period of motion. The quantities E and H are the characteristic mean absolute values of field strengths in the region of space occupied by the system of charges. Of course, it makes no sense to seek to define these quantities more precisely by relating them to a definite instant of time or to a definite point of space. The relations written have the meaning of rough estimates of the order of magnitude.

Further, let us find the order of magnitude of the spatial derivatives of the field in the same region of space. The fields \mathbf{E} and \mathbf{H} in real systems performing a quasistationary motion vary from point to point in general reasonably smoothly. If the mean dimensions of the system are denoted by L , then the order of magnitude of all spatial derivatives can be written in the form

$$\left| \frac{\partial \mathbf{E}}{\partial x} \right| \sim \left| \frac{\partial \mathbf{E}}{\partial y} \right| \sim \left| \frac{\partial \mathbf{E}}{\partial z} \right| \sim \frac{E}{L}.$$

In this rough estimate we disregard the field distribution in the system and the specific dependence on different coordinates.

The quasistationary condition is that the temporal variations of the fields should be sufficiently slow that in Maxwell's equations one can discard the terms containing the time derivatives which have small coefficients compared with those of the terms characterizing the spatial variation of the fields. For this the following inequalities must be fulfilled (to order of magnitude):

$$\left| \frac{\partial \mathbf{E}_i}{\partial x_k} \right| \gg \frac{1}{c} \left| \frac{\partial \mathbf{H}_i}{\partial t} \right|,$$

$$\left| \frac{\partial \mathbf{H}_i}{\partial x_k} \right| \gg \frac{1}{c} \left| \frac{\partial \mathbf{E}_i}{\partial t} \right|,$$

or

$$\frac{E}{L} \gg \frac{1}{c} \frac{H}{T}, \quad \frac{H}{L} \gg \frac{1}{c} \frac{E}{T}. \quad (19.4')$$

However, in this case the following equalities must be fulfilled:

$$\frac{\partial \mathbf{E}_i}{\partial x_k} \approx \frac{\partial \mathbf{E}_k}{\partial x_i}, \quad \frac{\partial \mathbf{H}_i}{\partial x_k} \approx \frac{\partial \mathbf{H}_k}{\partial x_i},$$

so that the differences between the spatial derivatives contained in the Maxwell equations compensate mutually for each other, and the derivatives with respect to time (with the coefficient $1/c$) turn out to be quantities of a higher degree of smallness.

Multiplying the inequalities (19.4') we arrive at the condition of quasistationarity

$$T \gg L/c. \quad (19.5)$$

The inequality (19.5) and the equivalent inequality

$$c \gg L/T \sim v, \quad (19.5')$$

where the quantity $v \sim L/T$ can be interpreted as the characteristic velocity

of the motion of charges in the system, have an obvious meaning. Later, in § 23, it will be shown that the velocity c contained in Maxwell's equations represents the velocity of propagation of the electromagnetic field in vacuum. For a quasistationary motion of charges their velocities must be small in comparison with the velocity of propagation of the field.

In this case the variation of the electromagnetic field in time is slow, so that the derivatives of the field with respect to time are of a higher degree of smallness than the spatial derivatives, and may be discarded.

Those electromagnetic fields for which the inequality (19.5) is valid and in which the displacement current can be regarded as small are called quasistationary fields.

For quasistationary fields the Maxwell equations assume the following form:

$$\nabla \times \mathbf{H} = \frac{4\pi}{c} \mathbf{j}, \quad (19.6)$$

$$\nabla \cdot \mathbf{H} = 0, \quad (19.7)$$

$$\nabla \times \mathbf{E} = 0, \quad (19.8)$$

$$\nabla \cdot \mathbf{E} = 4\pi\rho. \quad (19.9)$$

Thus, in the approximation of quasistationary fields the displacement current does not enter into the field equations. We have already seen that the absence of the displacement current corresponds to the solenoidal character of the current lines. Conversely, if current tubes are almost closed, while the motion of charges takes place in a limited volume and is almost periodic, then the displacement current must be very small in comparison with the current of the moving charges.

Maxwell's equations turn out to be resolved into equations for independent fields: the magnetic field due to currents and the electric field due to charges.

The charge density ρ in eq. (19.9) depends on time as a parameter. In the approximation of slowly moving charges the solution of the equations for the electric field leads to the following obvious result: at every instant of time the electric field is the same as the electrostatic field of the given configuration of charges.

The magnetic field of a system of charges performing a slow quasistationary motion can be found by integrating (19.6)–(19.7). We introduce the vector potential according to formula (10.1). Since the time dependence of the charge density ρ and the current \mathbf{j} can be disregarded, the strengths of the

magnetic and electric fields and, consequently, also the electromagnetic potentials do not depend on time. Hence the equations for the vector potential (10.6) and the Lorentz condition (10.5) assume the form

$$\nabla^2 \mathbf{A} = -\frac{4\pi}{c} \mathbf{j}, \quad (19.10)$$

$$\nabla \cdot \mathbf{A} = 0. \quad (19.11)$$

Eq. (19.10) represents a set of three scalar equations:

$$\nabla^2 A_i = -\frac{4\pi}{c} j_i \quad (i = x, y, z),$$

each of which is Poisson's equation.

We shall assume that all the components of the vector potential of a system of charges performing a slow quasistationary motion decrease at infinity not more slowly than $1/r$:

$$A_i \sim O(1/r) \rightarrow 0 \quad \text{as } r \rightarrow \infty. \quad (19.12)$$

Here the symbol O means that the discarded terms are of an order of magnitude less than $1/r$. The solution of eq. (19.10) satisfying the requirement (19.12) can be written according to formula (3.16) in the form

$$\mathbf{A}(\mathbf{r}) = \frac{1}{c} \int \frac{\mathbf{j}(\mathbf{r}') dV'}{R} = \frac{1}{c} \int \frac{\mathbf{j}(x', y', z') dx' dy' dz'}{\sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}}, \quad (19.13)$$

where $\mathbf{j}(\mathbf{r}')$ is the current density at a point \mathbf{r}' , and $R = |\mathbf{r} - \mathbf{r}'|$ is the distance from this point to the observation point \mathbf{r} at which the value of the vector potential is sought.

It is easy to see that the solution (19.13) of eq. (19.10) satisfies the additional condition (19.11). Indeed, we have

$$\nabla \cdot \mathbf{A} = \frac{1}{c} \nabla_{\mathbf{r}} \cdot \int \frac{\mathbf{j}(\mathbf{r}') dV'}{R} = \frac{1}{c} \int \nabla_{\mathbf{r}} \cdot \left(\frac{\mathbf{j}(\mathbf{r}')}{R} \right) dV',$$

where the divergence is taken with respect to the coordinates of the observation point. In view of the independence of the operations of differentiation with respect to the coordinates \mathbf{r} and integration with respect to the coor-

dinates \mathbf{r}' their order can be altered. The density of the current $\mathbf{j}(\mathbf{r}')$ could be brought outside the sign $\nabla_{\mathbf{r}'}$, but this serves no purpose. A substitution of the variable of differentiation according to a formula analogous to (I.17) gives

$$\nabla \cdot \mathbf{A} = -\frac{1}{c} \int \nabla_{\mathbf{r}'} \cdot \left(\frac{\mathbf{j}(\mathbf{r}')}{R} \right) dV' = -\frac{1}{c} \oint \frac{\mathbf{j} \cdot d\mathbf{S}}{R} = -\frac{1}{c} \oint \frac{j_n dS}{R} = 0,$$

by virtue of the condition (19.1).

Thus, formula (19.13) gives a solution of the problem satisfying all necessary conditions. Knowing the vector potential one can find the magnetic field

$$\mathbf{H} = \frac{1}{c} \nabla \times \int \frac{\mathbf{j} dV'}{R} = \frac{1}{c} \int \nabla_{\mathbf{r}'} \times \frac{\mathbf{j} dV'}{R}. \quad (19.14)$$

In differentiating with respect to the coordinates \mathbf{r} the density of the current $\mathbf{j}(\mathbf{r}')$ must be assumed to be constant. Then according to formula (I.43) we find

$$\nabla \times \frac{\mathbf{j}}{R} = \left(\nabla \frac{1}{R} \right) \times \mathbf{j} = \frac{\mathbf{j} \times \mathbf{R}}{R^3}.$$

Hence, finally,

$$\mathbf{H} = \frac{1}{c} \int \frac{\mathbf{j} \times \mathbf{R}}{R^3} dV'. \quad (19.15)$$

Formula (19.15) is called the Biot-Savart law. It gives, in principle, the solution of the problem posed. However, the calculation of the integral in formula (19.15) is rather complicated and can be carried through only for the simplest systems.

In the next paragraph formula (19.15) is applied to find the field of a single charge moving in vacuum. However, it is particularly important for the calculation of the fields of currents flowing in conductors. Hence further examples of calculations by means of the Biot-Savart law will be analyzed in § 17 of Part IV.

It should be stressed that all the results of this paragraph, in particular the Biot-Savart law, are of an approximate character. They are a consequence of the relations (19.5). Quasistationary fields are encountered particularly often in considering electromagnetic processes in material media. Hence we shall return to them in Part IV (§ 22), where the conditions under which a field can be assumed to be quasistationary will be discussed in more detail.

In conclusion we present a relation which we shall need later. The dimensions of that region of space in which the action of a magnetic field on a system is considered are very often sufficiently small that in the limits of this region the magnetic field can be assumed to be constant and uniform. Then the vector potential A of this constant uniform field can be written in the form

$$A = \frac{1}{2} H \times r. \quad (19.16)$$

One can verify the validity of this relation by a direct calculation according to formula (1.45).

§20. The field of a point charge undergoing a slow uniform motion

Let us consider a point charge e moving with a constant velocity v_0 (fig. 1.8). We assume the absolute value of the velocity $|v_0|$ to be small in comparison with the characteristic velocity c occurring in Maxwell's equations (see further §23).

The charge density ρ can be written in the form

$$\rho(r', t) = e\delta(r' - r_0(t)). \quad (20.1)$$

Here $r_0(t)$ is the coordinate of the charge at an instant t . Formula (20.1) means that at any given moment the charge is present at a point of space with the coordinate $r_0(t)$. Maxwell's equations for the electric field have the form

$$\nabla \times E = 0,$$

$$\nabla \cdot E = 4\pi e\delta(r - r_0(t)).$$

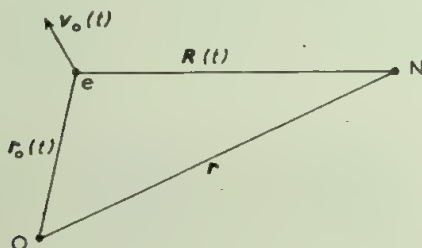


Fig. 1.8

Since the field is irrotational, a potential $\varphi(\mathbf{r}, t)$ can be introduced, so that

$$\mathbf{E} = - \nabla \varphi.$$

The potential satisfies the equation

$$\nabla^2 \varphi = - 4\pi e \delta(\mathbf{r} - \mathbf{r}_0(t)).$$

The solution of the equation can, according to (15.1), be written in the form

$$\begin{aligned} \varphi &= \int \frac{\rho dV'}{|\mathbf{r} - \mathbf{r}'|} = e \int \frac{\delta(\mathbf{r}' - \mathbf{r}_0(t))}{|\mathbf{r} - \mathbf{r}'|} dV' = \\ &= \frac{e}{|\mathbf{r} - \mathbf{r}_0(t)|} = \frac{e}{R(t)}. \end{aligned} \quad (20.2)$$

Here $\mathbf{R}(t)$ stands for a vector connecting the point of observation N with the instantaneous charge coordinate $\mathbf{r}_0(t)$.

From formula (20.2) it follows that the electric field of the moving charge is formally the same as that of a charge at rest, but, instead of the fixed distance from the given point of observation to the charge, the time-dependent distance $R(t)$ figures in (20.2). The field of the charge is obviously given by the formula

$$\mathbf{E} = \frac{e \mathbf{R}(t)}{R^3(t)}. \quad (20.3)$$

Since the charge is moving uniformly, its position in space can be written in the form

$$\mathbf{r}_0 = \mathbf{v}_0 t.$$

Hence the electric field strength at a point will depend on time according to the law

$$\mathbf{E}(\mathbf{r}, t) = \frac{e(\mathbf{r} - \mathbf{v}_0 t)}{|\mathbf{r} - \mathbf{v}_0 t|^3}. \quad (20.4)$$

It is obvious that at a certain point \mathbf{r} at time t the field strength will be the same as that at a point with the coordinate $\mathbf{r} + \mathbf{v}_0$ at time $t + 1$. Indeed,

$$\mathbf{E}(\mathbf{r} + \mathbf{v}_0, t + 1) = \frac{e[\mathbf{r} + \mathbf{v}_0 - \mathbf{v}_0(t + 1)]}{|\mathbf{r} + \mathbf{v}_0 - \mathbf{v}_0(t + 1)|^3} = \mathbf{E}(\mathbf{r}, t).$$

This means that the point with given value of field strength moves in space together with the charge. The field retains spherical symmetry with respect to the instantaneous position of the charge. Later we shall compare formula (20.4) with the corresponding expression for the field produced by a charge moving with a velocity $|\mathbf{v}_0| \approx c$.

Let us now pass on to the determination of the magnetic field. It can be found by means of formula (19.15), in which the current density for unit charge can be written in the form

$$\mathbf{j} = e \mathbf{v}_0 \delta(\mathbf{r} - \mathbf{r}_0).$$

Substituting the value of \mathbf{j} into (19.15), we find

$$\mathbf{H} = \frac{e}{c} \int \frac{[\mathbf{v}_0 \times (\mathbf{r} - \mathbf{r}')] \delta(\mathbf{r}' - \mathbf{r}_0)}{|\mathbf{r} - \mathbf{r}'|^3} dV' = \frac{e}{c} \frac{\mathbf{v}_0 \times \mathbf{R}}{R^3} = \frac{1}{c} \mathbf{v}_0 \times \mathbf{E}. \quad (20.5)$$

Thus, the magnetic field turns out to be perpendicular to the electric field and to the velocity of the charge. The absolute value $|\mathbf{H}| \sim (v_0/c) |\mathbf{E}|$, i.e. the absolute value of the magnetic field, is small in comparison with the electric field by the ratio v_0/c . Differentiating according to formula (1.45) one can see that the vector \mathbf{H} satisfies eq. (19.7).

In §9 we have discussed, without calculation, the problem of finding the field of a moving charge. We have shown that if the integration surface S_2 (fig. 4) passes through the point at which the moving charge is present at a given instant, the magnetic field is connected with the charge current \mathbf{j} . This picture corresponds to the calculation we have performed.

Formula (20.5) has been obtained as a result of solving eqs. (19.6) and (19.7), in which the displacement current was absent. However, it is possible to find the value of the magnetic field from the displacement current on an arbitrary surface S_1 , without making use of formula (19.15).

We write the equations for the magnetic field in the form

$$\nabla \times \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} e \mathbf{v}_0 \delta(\mathbf{r} - \mathbf{r}_0), \quad (20.6)$$

$$\nabla \cdot \mathbf{H} = 0. \quad (20.7)$$

The electric field \mathbf{E} of the uniformly moving charge depends on time according to formula (20.4). Differentiating \mathbf{E} with respect to time, we find

$$\frac{\partial \mathbf{E}}{\partial t} = - \left(\frac{\partial \mathbf{E}}{\partial x} v_x + \frac{\partial \mathbf{E}}{\partial y} v_y + \frac{\partial \mathbf{E}}{\partial z} v_z \right) = -(\mathbf{v}_0 \cdot \nabla) \mathbf{E}.$$

From formula (1.45) we have

$$\begin{aligned} \nabla \times (\mathbf{v}_0 \times \mathbf{E}) &= (\mathbf{E} \cdot \nabla) \mathbf{v}_0 - (\mathbf{v}_0 \cdot \nabla) \mathbf{E} + \mathbf{v}_0 (\nabla \cdot \mathbf{E}) - \mathbf{E} (\nabla \cdot \mathbf{v}_0) = \\ &= -(\mathbf{v}_0 \cdot \nabla) \mathbf{E} + \mathbf{v}_0 (\nabla \cdot \mathbf{E}), \end{aligned}$$

since \mathbf{v}_0 is a constant vector. Whence we find

$$\begin{aligned} \frac{\partial \mathbf{E}}{\partial t} &= -(\mathbf{v}_0 \cdot \nabla) \mathbf{E} = \nabla \times (\mathbf{v}_0 \times \mathbf{E}) - \mathbf{v}_0 (\nabla \cdot \mathbf{E}) = \\ &= \nabla \times (\mathbf{v}_0 \times \mathbf{E}) - 4\pi \mathbf{v}_0 e \delta(\mathbf{r} - \mathbf{r}_0), \end{aligned} \quad (20.8)$$

by virtue of (20.1).

Substituting (20.8) into (20.6), we have

$$\nabla \times \mathbf{H} = \frac{1}{c} \nabla \times (\mathbf{v}_0 \times \mathbf{E}). \quad (20.9)$$

Expression (20.5) provides the solution of (20.9) satisfying (20.7). Thus we see that the two methods of calculation lead to the same result, as was to be expected.

§ 21. The field of a system of charges undergoing quasistationary motion at a large distance from the system

Let us assume that a certain set of charges performs a slow and quasistationary motion in a limited region of space. The electromagnetic field of this system at a distance large in comparison with the dimensions of the system (dimensions of the region of motion) is often of great interest. Here, as well as in the case of electrostatics, the general formula for the vector potential (19.13) allows us to make a substantial simplification.

We make use of the expansion (15.8) writing

$$\frac{1}{R} \approx \frac{1}{r} - \left(\mathbf{r}' \cdot \nabla \frac{1}{r} \right).$$

Substituting $1/R$ into (19.13), we find

$$\mathbf{A} = \frac{1}{c} \frac{1}{r} \int \mathbf{j} dV' - \frac{1}{c} \int \mathbf{j} \left(\mathbf{r}' \cdot \nabla \frac{1}{r} \right) dV'.$$

By virtue of (19.4) for a system performing stationary motion the first integral is equal to zero. Hence

$$\mathbf{A} = -\frac{1}{c} \int \mathbf{j} \left(\mathbf{r}' \cdot \nabla \frac{1}{r} \right) dV'. \quad (21.1)$$

We transform the integrand in (21.1) by means of the identity

$$\begin{aligned} \mathbf{j} \left(\mathbf{r}' \cdot \nabla \frac{1}{r} \right) &\equiv \frac{1}{2} \left\{ \mathbf{j} \left(\mathbf{r}' \cdot \nabla \frac{1}{r} \right) - \mathbf{r}' \left(\mathbf{j} \cdot \nabla \frac{1}{r} \right) \right\} + \\ &+ \frac{1}{2} \left\{ \mathbf{j} \left(\mathbf{r}' \cdot \nabla \frac{1}{r} \right) + \mathbf{r}' \left(\mathbf{j} \cdot \nabla \frac{1}{r} \right) \right\}. \end{aligned}$$

The first bracketed pair of terms can be presented in the form of the triple vector product:

$$\left\{ \mathbf{j} \left(\mathbf{r}' \cdot \nabla \frac{1}{r} \right) - \mathbf{r}' \left(\mathbf{j} \cdot \nabla \frac{1}{r} \right) \right\} = (\mathbf{r}' \times \mathbf{j}) \times \nabla \frac{1}{r} = -\frac{(\mathbf{r}' \times \mathbf{j}) \times \mathbf{r}}{r^3}.$$

Thus the vector potential can be written in the form

$$\mathbf{A} = \frac{1}{2c} \int \frac{(\mathbf{r}' \times \mathbf{j}) \times \mathbf{r}}{r^3} dV' - \frac{1}{2c} \int \left\{ \mathbf{j} \left(\mathbf{r}' \cdot \nabla \frac{1}{r} \right) + \mathbf{r}' \left(\mathbf{j} \cdot \nabla \frac{1}{r} \right) \right\} dV'.$$

We go on to the calculation of the second integral. Introducing current tubes, according to the general theory of § 19, we can write

$$\mathbf{j} dV' = dI d\mathbf{l} = dI d\mathbf{r}'.$$

Indeed, the change of the charge position $d\mathbf{r}'$ in its motion along the current tube is identical with $d\mathbf{l}$. Hence

$$\begin{aligned}
 I &= \int \left\{ \mathbf{j} \left(\mathbf{r}' \cdot \nabla \frac{1}{r} \right) + \mathbf{r}' \left(\mathbf{j} \cdot \nabla \frac{1}{r} \right) \right\} dV' = \\
 &= \int dI \oint \left\{ d\mathbf{r}' \left(\mathbf{r}' \cdot \nabla \frac{1}{r} \right) + \mathbf{r}' \left(d\mathbf{r}' \cdot \nabla \frac{1}{r} \right) \right\} = \\
 &= \int dI \oint d \left(\mathbf{r}' \left(\mathbf{r}' \cdot \nabla \frac{1}{r} \right) \right) = 0,
 \end{aligned}$$

since the integral over a closed contour of the total differential is always equal to zero.

Thus, finally,

$$\mathbf{A} = \frac{1}{2cr^3} \int (\mathbf{r}' \times \mathbf{j}) \times \mathbf{r} dV' = \frac{1}{2cr^3} \int (\mathbf{r}' \times \mathbf{j}) dV' \times \mathbf{r}. \quad (21.2)$$

We introduce the notation

$$\mathbf{M} = \frac{1}{2c} \int \mathbf{r}' \times \mathbf{j} dV'. \quad (21.3)$$

The quantity \mathbf{M} is called the magnetic moment of the system of charges. It depends solely on the properties of the charge system — the current density distribution and the geometry of the system. We shall see further that \mathbf{M} is indeed, to a certain degree, an analogue of the dipole moment of a system of charges at rest.

The vector potential at a large distance from the system assumes the form

$$\mathbf{A} = \frac{\mathbf{M} \times \mathbf{r}}{r^3}. \quad (21.4)$$

The magnetic field is expressed, according to (1.45), by the formula

$$\mathbf{H} = \nabla \times \mathbf{A} = \nabla \times \frac{\mathbf{M} \times \mathbf{r}}{r^3} = \mathbf{M} \left(\nabla \cdot \frac{\mathbf{r}}{r^3} \right) - (\mathbf{M} \cdot \nabla) \frac{\mathbf{r}}{r^3}.$$

Since the differentiation is carried out with respect to the coordinates of the observation point, the magnetic moment of the system is constant in the differentiation.

According to (1.42) we find

$$\nabla \cdot \frac{\mathbf{r}}{r^3} = \mathbf{r} \cdot \nabla \frac{1}{r^3} + \frac{1}{r^3} \nabla \cdot \mathbf{r} = -\frac{3}{r^3} + \frac{3}{r^3} = 0.$$

Further,

$$(\mathbf{M} \cdot \nabla) \frac{\mathbf{r}}{r^3} = \frac{1}{r^3} (\mathbf{M} \cdot \nabla) \mathbf{r} + \mathbf{r} \left(\mathbf{M} \cdot \nabla \frac{1}{r^3} \right) = \frac{\mathbf{M}}{r^3} - \frac{3\mathbf{r}(\mathbf{M} \cdot \mathbf{r})}{r^5}.$$

Hence

$$\mathbf{H} = \frac{3\mathbf{r}(\mathbf{M} \cdot \mathbf{r}) - r^2 \mathbf{M}}{r^3}. \quad (21.5)$$

We see that the magnetic field at a large distance from a system of charges in slow and quasistationary motion, is expressed by the same formula as the electrostatic field of a system of charges at rest. The difference consists of the fact that, instead of the dipole moment \mathbf{d} in (15.18), the magnetic moment \mathbf{M} of the system occurs in formula (21.5).

§22. The magnetic moment

We shall consider in somewhat more detail the properties of the magnetic moment of a system. First of all it is easy to verify the fact that the value of the magnetic moment does not depend on the choice of the origin. Shifting the origin by a constant vector \mathbf{a} , i.e. assuming $\mathbf{r}' = \mathbf{r}'' + \mathbf{a}$, we find

$$\mathbf{M} = \frac{1}{2c} \int \mathbf{r}' \times \mathbf{j} dV' = \frac{1}{2c} \int \mathbf{r}'' \times \mathbf{j} dV' + \frac{1}{2c} \int \mathbf{a} \times \mathbf{j} dV'.$$

Rewriting the additional term in the form

$$\int \mathbf{a} \times \mathbf{j} dV' = \mathbf{a} \times \int \mathbf{j} dV',$$

we see by virtue of (19.4) that it is equal to zero.

Thus, the magnetic moment, like the dipole moment of a neutral system of charges, represents a quantity which depends only on the physical properties of the system but not on the choice of the origin.

Consider the expression of the magnetic moment in the case when charges

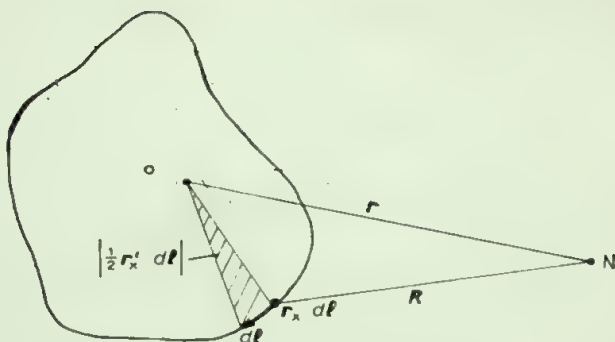


Fig. I.9

are moving along a thread or a tube (fig. I.9). Making use of (19.3), we find

$$\mathbf{M} = \frac{1}{2c} \int dI \mathbf{r}' \times d\mathbf{l} = \frac{I}{c} \frac{1}{2} \int \mathbf{r}' \times d\mathbf{l}. \quad (22.1)$$

It is easy to see that the quantity $\frac{1}{2} \mathbf{r}' \times d\mathbf{l}$ represents an area vector. The integral $\mathbf{S} = \int \frac{1}{2} \mathbf{r}' \times d\mathbf{l}$ represents the area of the lateral surface of a cone resting on the path of the current. In the particular case of a flat closed path, one can choose as \mathbf{S} the vector of the normal to the plane of the path multiplied by the area of the plane. Then

$$\mathbf{M} = \frac{IS}{c}. \quad (22.2)$$

Formula (22.2) yields to an obvious interpretation: every closed current (for example, one or several charged particles moving along closed trajectories) possesses a magnetic moment proportional to the value of the current. In this connection we note that every atom with electrons rotating in an orbit is an elementary magnet (see Part IV). We now rewrite the expression for the magnetic moment by specifying the current density in terms of the velocity of motion of the charges:

$$\mathbf{M} = \frac{1}{2c} \int \mathbf{r}' \times \rho \mathbf{v} dV' = \frac{1}{2c} \sum \mathbf{r}'_i \times e \mathbf{v}_i, \quad (22.3)$$

where the summation is carried out over all the charges in the system.

Consider the important case of a system consisting of identical or different particles having the same value of the ratio of the charge e to the mass m . Introducing this relation into (22.3), we find

$$\mathbf{M} = \frac{1}{2c} \sum \frac{e_i}{m_i} \mathbf{r}'_i \times m_i \mathbf{v}_i = \frac{e}{2mc} \sum \mathbf{r}'_i \times \mathbf{p}_i = \frac{e}{2mc} \mathbf{L}, \quad (22.4)$$

where \mathbf{L} is the angular momentum of the system.

Formula (22.4) shows that for a system of particles with a fixed value of e/m there exists a direct proportionality between the magnetic moment and the angular momentum of the system.

The proportionality between \mathbf{M} and \mathbf{L} holds also for a system consisting of two particles with an arbitrary ratio e/m . Indeed, in such a system

$$\mathbf{M} = \frac{1}{2c} [e_1 \mathbf{r}'_1 \times \mathbf{v}_1 + e_2 \mathbf{r}'_2 \times \mathbf{v}_2].$$

Locating the origin at the centre of mass, i.e. assuming

$$m_1 \mathbf{r}'_1 + m_2 \mathbf{r}'_2 = 0,$$

and introducing the relative velocity

$$\mathbf{v}_{\text{rel}} = \dot{\mathbf{r}}'_2 - \dot{\mathbf{r}}'_1 = \dot{\mathbf{r}}'_{\text{rel}},$$

we find

$$\mathbf{v}_2 = \mathbf{v}_{\text{rel}} \frac{m_1}{m_1 + m_2}, \quad \mathbf{v}_1 = -\mathbf{v}_{\text{rel}} \frac{m_2}{m_1 + m_2},$$

$$\mathbf{M} = \frac{\mu^2}{2c} \left[\frac{e_1}{m_1^2} \mathbf{r}'_{\text{rel}} \times \mathbf{v}_{\text{rel}} + \frac{e_2}{m_2^2} \mathbf{r}'_{\text{rel}} \times \mathbf{v}_{\text{rel}} \right] = \frac{1}{2c} \left(\frac{e_1}{m_1^2} + \frac{e_2}{m_2^2} \right) \mu \mathbf{L},$$

where

$$\mathbf{L} = \mathbf{r}'_{\text{rel}} \times \mathbf{p}_{\text{rel}} = \mathbf{r}'_{\text{rel}} \times \mu \mathbf{v}_{\text{rel}}$$

is the angular momentum of the relative motion, and $\mu = m_1 m_2 / (m_1 + m_2)$ is the reduced mass.

The Electromagnetic Field of Arbitrarily Moving Charges

§ 23. The electromagnetic field of a system of arbitrarily moving charges

Let us consider a system of charges performing an arbitrary motion in a certain volume V' . The distribution and motion of charges in this volume will be characterized by the charge density $\rho(\mathbf{r}, t)$ and the current density $\mathbf{j}(\mathbf{r}, t)$, varying in space and time. We shall assume that the functions $\rho(\mathbf{r}, t)$ and $\mathbf{j}(\mathbf{r}, t)$ are known for all times (i.e. for $-\infty < t < \infty$).

Equations for the space-dependent and time-dependent electromagnetic potentials are of the form

$$\nabla^2 \varphi(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 \varphi(\mathbf{r}, t)}{\partial t^2} = -4\pi\rho(\mathbf{r}, t), \quad (23.1)$$

$$\nabla^2 \mathbf{A}(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}(\mathbf{r}, t)}{\partial t^2} = -\frac{4\pi}{c} \mathbf{j}(\mathbf{r}, t), \quad (23.2)$$

$$\nabla \cdot \mathbf{A}(\mathbf{r}, t) + \frac{1}{c} \frac{\partial \varphi(\mathbf{r}, t)}{\partial t} = 0. \quad (23.3)$$

The system of eqs. (23.1)–(23.3) represents a system of linear partial differential equations. As is known from the theory of partial differential equations

a single-value solution of the problem — in this case finding the particular distribution of the electromagnetic field in space and time as a function of the values of the known functions $\rho(\mathbf{r}, t)$ and $\mathbf{j}(\mathbf{r}, t)$ — requires that some further conditions, called boundary and initial conditions, should be given.

As a rule, the problem of finding the electromagnetic field is defined as follows: before a certain time $t = 0$ (i.e. for all $t < 0$) the charges of the system were at rest; at time $t = 0$ the charges are set in an arbitrary motion. A change or perturbation then arises in the electromagnetic field. We shall assume that eqs. (23.1)–(23.3) involve the vector and scalar potentials of just this perturbed field. The functions $\rho(\mathbf{r}, t)$ and $\mathbf{j}(\mathbf{r}, t)$, responsible for the perturbation of the field for $t > 0$, are considered as known. For $t \leq 0$ one sets

$$\rho(\mathbf{r}, 0) = 0, \quad \mathbf{j}(\mathbf{r}, 0) = 0.$$

Correspondingly, at the initial time $t = 0$ the vectors of the electric and magnetic field are equal to zero: $\mathbf{E}(\mathbf{r}, 0) = \mathbf{H}(\mathbf{r}, 0) = 0$. Then the initial conditions for the potentials read

$$\begin{aligned} A(\mathbf{r}, t) \Big|_{t=0} &= 0, \\ \frac{\partial A(\mathbf{r}, t)}{\partial t} \Big|_{t=0} &= 0, \\ \varphi(\mathbf{r}, t) \Big|_{t=0} &= 0. \end{aligned} \tag{23.4}$$

Indeed, if the conditions (23.4) are satisfied, then from the definition of the field vectors it is seen directly that the field vectors are zero. Thus, the initial condition (23.4) is completely equivalent to the requirement that the densities of the charge and current should be equal to zero for $t = 0$ and, in general, different from zero for $t > 0$.

The requirements that at a large distance from the volume V' the field potentials should decrease not more slowly than the law

$$\begin{aligned} \varphi &\sim O(1/r) \quad \text{as } r \rightarrow \infty, \quad 0 \leq t < \infty, \\ |A| &\sim O(1/r) \quad \text{as } r \rightarrow \infty, \quad 0 \leq t < \infty, \end{aligned} \tag{23.5}$$

i.e. not more slowly than the function $1/r$, will serve as the boundary conditions.

In solving the system of field equations we shall, in this paragraph, make use of a simple but not exact method, based on the use of the principle of superposition. In the next paragraph we shall present a method which is more consistent from the mathematical standpoint and which leads to the same results.

We divide the whole system into a set of arbitrarily small charges $\delta e_i = \rho(\mathbf{r}, t) \delta V_i$, where δV_i is an arbitrarily small volume within the region V' . We seek the potential of the field produced by the charge δe_i at a certain point of observation N , assuming that there are no other charges in the space. The total field, on the basis of the principle of superposition, represents the sum of fields produced by all the charges δe_i constituting the system.

It should be emphasized that the charge conservation law does not allow the existence of a time-dependent and solitary charge δe_i . In reality a change (for example, an increase) of the charge δe_i in time assumes a simultaneous decrease of a charge δe_k in another element of the volume, so that the total charge of the system is conserved. However, in finding the potentials of the field produced by the charge δe_i we shall not formally take into account the existence of other charges. This apparent violation of the charge conservation law will not affect the final solution, in which the summation over all charges of the system will be performed.

Let us first of all find the solution of the system of equations for the potentials of the field produced in the entire space outside the small volume δV_i by the charge δe_i .

At all points of this space outside the volume δV_i the charge density, according to our assumption, is equal to zero. Hence outside the volume δV_i the equations for the electromagnetic field potentials assume the form

$$\begin{aligned}\nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} &= 0, \\ \nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} &= 0, \\ \nabla \cdot \mathbf{A} + \frac{1}{c} \frac{\partial \varphi}{\partial t} &= 0.\end{aligned}\tag{23.6}$$

We introduce spherical coordinates with the origin located in the volume δV_i . The field outside the volume δV_i possesses spherical symmetry, so that the field potentials can depend only on the distance from the volume δV_i — the radius vector \mathbf{r} — and time. From the expression for the Laplacian in

spherical coordinates, we have

$$\frac{1}{r} \frac{\partial^2}{\partial r^2} (r\varphi) - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = 0, \quad (23.7)$$

$$\frac{1}{r} \frac{\partial^2}{\partial r^2} (rA) - \frac{1}{c^2} \frac{\partial^2 A}{\partial t^2} = 0. \quad (23.8)$$

We see that the scalar potential and all the components of the vector potential are determined by an equation of one type:

$$\frac{1}{r} \frac{\partial^2}{\partial r^2} (rf) - \frac{1}{c^2} \frac{\partial^2 f}{\partial t^2} = 0. \quad (23.9)$$

For reasons which will be clear from what follows, an equation of the type (23.9) is called a wave equation. The integration of the wave equation can be carried out most simply by D'Alembert's method. D'Alembert's method consists, roughly speaking, of the reduction of a partial differential equation of the type (23.9) to an equation with a mixed second derivative (for a more strict presentation of D'Alembert's method we refer the reader to mathematical texts *).

We rewrite eq. (23.9) in the form

$$\frac{\partial^2}{\partial r^2} (rf) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} (rf) = 0, \quad (23.10)$$

and introduce a new unknown function

$$\psi = rf. \quad (23.11)$$

This is always possible since $r \neq 0$ outside the volume δV_i . We then have

$$\frac{\partial^2 \psi}{\partial r^2} - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = 0. \quad (23.12)$$

* See, for example, A. N. Tikhonov and A. A. Samarskii, *Partial differential equations of mathematical physics* (Holden Day, San Francisco, 1964) Ch. II.

Now we change over to new variables in eq. (23.12):

$$\xi = t - \frac{r}{c}, \quad (23.13)$$

$$\eta = t + \frac{r}{c}. \quad (23.14)$$

Whence

$$t = \frac{1}{2}(\xi + \eta), \quad r = \frac{1}{2}(\eta - \xi)c,$$

so that

$$\frac{\partial}{\partial \xi} = \frac{\partial}{\partial r} \frac{\partial r}{\partial \xi} + \frac{\partial}{\partial t} \frac{\partial t}{\partial \xi} = \frac{1}{2}c \left(-\frac{\partial}{\partial r} + \frac{1}{c} \frac{\partial}{\partial t} \right),$$

$$\frac{\partial}{\partial \eta} = \frac{\partial}{\partial r} \frac{\partial r}{\partial \eta} + \frac{\partial}{\partial t} \frac{\partial t}{\partial \eta} = \frac{1}{2}c \left(\frac{\partial}{\partial r} + \frac{1}{c} \frac{\partial}{\partial t} \right).$$

Further

$$\frac{\partial^2}{\partial r^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} = \left(\frac{\partial}{\partial r} + \frac{1}{c} \frac{\partial}{\partial t} \right) \left(\frac{\partial}{\partial r} - \frac{1}{c} \frac{\partial}{\partial t} \right) = -\frac{4}{c^2} \frac{\partial}{\partial \xi} \frac{\partial}{\partial \eta} = -\frac{4}{c^2} \frac{\partial^2}{\partial \xi \partial \eta}.$$

Eq. (23.12) in new variables has the form

$$\frac{\partial^2 \psi}{\partial \xi \partial \eta} = 0. \quad (23.15)$$

This equation, containing only the mixed derivative, is integrated directly. It is obvious that the equation is satisfied by any functions $\psi_1(\eta)$ and $\psi_2(\xi)$ of one variable, ξ or η . Hence the general solution of eq. (23.15) can be written in the form

$$\psi(\xi, \eta) = \psi_1(\xi) + \psi_2(\eta), \quad (23.16)$$

where ψ_1 and ψ_2 are arbitrary functions of one variable ξ and η respectively. Returning to the previous variables r and t , we obtain

$$\psi(r, t) = \psi_1 \left(t - \frac{r}{c} \right) + \psi_2 \left(t + \frac{r}{c} \right). \quad (23.17)$$

This solution has a simple meaning. The value of the function ψ_1 at the point $(r+c)$ at time $(t+1)$ is the same as at the point r at time t . This means that $\psi_1(t-(r/c))$ describes a process which is periodic in space and time, i.e. a wave process. The wave propagates in the direction of increasing values of the distance r from the origin with a velocity equal to c . Similarly $\psi_2(t+(r/c))$ describes a wave propagating from large r 's to smaller ones in the direction toward the origin. For the function f we have

$$f = \frac{\psi_1\left(t - \frac{r}{c}\right)}{r} + \frac{\psi_2\left(t + \frac{r}{c}\right)}{r} \quad (23.18)$$

Formula (23.18), giving the general solution of eq. (23.9), represents the superposition of two waves, the one diverging from the origin (the first term) and the other converging to the origin (the second term). The surfaces of the spheres $r = \text{const}$ are the surfaces of constant value of the function f or equal-phase surfaces. Since equal-phase surfaces are spherical surfaces, it is said that the wave process described by the function f is the combination of a diverging and a converging spherical wave. The scalar potential φ and all components of the vector potential \mathbf{A} can be presented in the form of formula (23.18).

In order to make clear the meaning of the solutions obtained, we shall consider one of the particular solutions, for instance the diverging spherical wave. For concreteness we write the expression for the scalar potential:

$$\varphi(r, t) = \frac{\varphi_1\left(t - \frac{r}{c}\right)}{r} \quad (23.19)$$

For an arbitrary form of the function φ_1 formula (23.19) gives the particular solution of eq. (23.1) in the region of space outside the volume δV_i . We now require that (23.19) should pass over continuously into the solution of eq. (23.1) near the volume δV_i , i.e. near the locus of the charge $\delta e_i(t)$. If in eq. (23.1) a formal transition is made by assuming $c \rightarrow \infty$ (the meaning of such a transition will become clear from what follows), then it will obviously turn into the equation for the electrostatic potential, the solution of which is

$$\Delta \varphi_i = \frac{\delta e_i(t)}{r} = \frac{\rho(r, t) \delta V_i}{r} \quad (23.20)$$

Writing (23.19) in the form

$$d\varphi_i(r, t) = \frac{\rho \left(t - \frac{r}{c} \right) \delta V_i}{r}, \quad (23.21)$$

we arrive at the potential of the field produced by the charge δe_i , which satisfies eq. (23.7) outside the volume δV_i and turns into (23.20) near the origin.

Formula (23.21) shows that the field potential at the observation point a distance r from the origin at time t is determined by the value of the charge at the preceding instant $\tau = t - (r/c)$. The potential (23.21) is therefore called retarded potential, while the quantity r/c is called the delay time. The delay time represents the time interval during which the electromagnetic field, propagating with the velocity c , traverses the path r .

Introducing the origin at point O, located in the volume V' , and integrating the expression (23.21) over all charges of the system, we arrive at the following expression for the potential of the field produced at the point of observation N:

$$\begin{aligned} \varphi(r, t) &= \int \frac{\rho \left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c} \right) dV'}{|\mathbf{r} - \mathbf{r}'|} = \\ &= \int \frac{\rho \left(\mathbf{r}', t - \frac{R}{c} \right) dV'}{R} = \int \frac{\rho(\mathbf{r}', \tau) dV'}{R}, \end{aligned} \quad (23.22)$$

where $\tau = t - (R/c)$, $\mathbf{R} = \mathbf{r} - \mathbf{r}'$.

According to (23.22), to obtain the potential at the point of observation N, it is necessary as in electrostatics, to evaluate the integral of the quantity ρ/R over the whole volume of the system. In this case, however, the value of the charge density is taken at time $\tau = t - |\mathbf{r} - \mathbf{r}'|/c$, where the delay time $|\mathbf{r} - \mathbf{r}'|/c$ is determined by the distance from every point \mathbf{r}' to the observation point.

Similarly, for the vector potential the particular solution of eq. (23.2) can be written in the form

$$\begin{aligned} \mathbf{A} &= \frac{1}{c} \int \frac{\mathbf{j} \left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c} \right) dV'}{R} = \\ &= \frac{1}{c} \int \frac{\mathbf{j} \left(\mathbf{r}', t - \frac{R}{c} \right) dV'}{R} = \frac{1}{c} \int \frac{\mathbf{j}(\mathbf{r}', \tau) dV'}{R}. \end{aligned} \quad (23.23)$$

In addition to the solutions of the equations for potentials in the form of retarded potentials, one can also write other particular solutions corresponding to the function ψ_2 in the general solution (23.18):

$$\varphi^*(\mathbf{r}, t) = \int \frac{\rho\left(\mathbf{r}', t + \frac{R}{c}\right) dV'}{R}, \quad (23.24)$$

$$\mathbf{A}^*(\mathbf{r}, t) = \frac{1}{c} \int \frac{\mathbf{j}\left(\mathbf{r}', t + \frac{R}{c}\right) dV'}{R} \quad (23.25)$$

In formulae (23.24) and (23.25) the values of the functions ρ and \mathbf{j} , determining the potentials at the point \mathbf{r} at time t , are taken at the instant $\tau^* = t + (R/c)$. This means that the potentials at the instant t depend on that charge density which will appear in the point \mathbf{r} after the time interval R/c . The potentials (23.24) and (23.25) are called advanced potentials.

From retarded and advanced potentials one can make arbitrary linear combinations of the form $\alpha_1\varphi + \alpha_2\varphi^*$, $\beta_1\mathbf{A} + \beta_2\mathbf{A}^*$, which also satisfy the field equations. The general solution of the equations for the potentials is obtained from the particular solutions and general solutions found for eqs. (23.7) and (23.8).

The appearance of retarded and advanced potentials as equally valid solutions of the field equations is quite natural. Like the equations of mechanics, the electrodynamical equations are symmetric with respect to the future and the past. They do not change when t is substituted for $(-t)$ and, hence, must have a general solution invariant under the change of the sign of time.

The choice of coefficients in these linear combinations is determined by giving the above mentioned supplementary conditions characterizing the behaviour of potentials at infinity. In order that these conditions may be fulfilled it is necessary to reject the solution corresponding to the advanced potentials. Indeed, let us consider at the moment $t = 0$ a certain sphere of radius R_1 outside the volume V' . According to the boundary condition (23.5), the potential $\varphi(R_1, 0) \sim O(1/R_1)$. Correspondingly, the retarded potential

$$\varphi(R_1, 0) = \int \frac{\rho(\mathbf{r}', -R_1/c)}{R} dV',$$

since according to the condition the charge density is equal to zero for $t < 0$. On the contrary, the advanced potential

$$\varphi^*(R_1, 0) = \int \frac{\rho(\mathbf{r}', R_1/c)}{R} dV' \neq O(1/R_1).$$

In the last formula there corresponds to time $t = 0$, a non-zero value of the argument of the function $\rho(\mathbf{r}', \tau)$, because the charge density $\rho(\mathbf{r}', R_1/c)$ is not equal to zero.

It should be noted that the properties of the charge density $\rho(\mathbf{r}', 0) = 0$, $\rho(\mathbf{r}', t - (R/c)) \neq 0$, used here, correspond to the choice of initial conditions in the form of (23.4). We see that, for such properties of the density ρ , the advanced potential does not satisfy the condition (23.5) determining its behaviour at infinity. Namely, the advanced potential decreases more slowly than the function $1/r$. In order to obtain a solution of the wave equation which satisfies the system of initial and boundary conditions, one has to assume that $\alpha_2 = \beta_2 = 0$ and retain only the solution of the field equations in the form of retarded potentials. Uncomplicated, but somewhat cumbersome calculations allow one to verify the fact that the retarded potentials given by formulae (23.24) and (23.25) satisfy the Lorentz condition (23.3) for $\alpha_1 = \beta_1 = 1$.

Thus, we arrive at the very important conclusion: a system of charges, which at the moment $t = 0$ begin to perform a non-steady state motion, produce in the surrounding space an electromagnetic field whose potentials have the character of retarded potentials. The field potentials have the form of spherical waves originating from the system and propagating in vacuum with the velocity c .

We shall say that a system of non-steady state moving charges emits electromagnetic waves, and call it briefly the emitter.

The solution of the electromagnetic field equations in the form of retarded potentials is of great significance. It corresponds to a definite concept about the character of causal relationship, which differs from the concepts of classical mechanics.

As is well-known, all the propositions of classical mechanics are in accordance with the Newtonian concept of action at a distance. In classical mechanics it is assumed that the acceleration of a material point at a given instant is completely determined by the force acting on it at that instant. The force acting on a given material point depends in turn on the position of other material points, located at a finite distance from the observed point. If the position of any of the material points is changed at a certain instant t_0 , then the magnitude of the force will be changed also at the same instant. In other words, the velocity of propagation of the interaction in space is considered in classical mechanics to be infinitely large.

In the theory of the electromagnetic field the situation is radically altered. If the position of the charges located at a distance r from the point of observation is changed, then the potential at the latter point will change only after the time $\tau = r/c$. This time is needed for the perturbation of the electromagnetic field, moving from point to point with a finite velocity equal to the velocity of light c , to traverse in space the path r . The perturbation is transferred from one point of the field to the neighbouring one. The space in which the propagation of electromagnetic perturbations occurs is no more the empty "nothing" of classical mechanics, but is considered to be filled by a real electromagnetic field endowed with definite physical properties. Thus, in the theory of the electromagnetic field the infinitely large velocity of propagation of interactions and the long-range action of classical mechanics are replaced by a finite velocity of propagation of interaction and the concept of short-range action. The cause (change of the field) and the effect (motion of the test charge at the observation point) refer to one location and one instant of time. In the next part of the book the standpoint of field theory will receive further confirmation and extension.

It should be noted that, because the velocity c of propagation of the electromagnetic perturbations is very large, it can often in practice be considered as infinitely large. The concepts of classical mechanics are by this fact not simply discarded as untrue, but are retained as approximate ones, having a well defined region of application.

§ 24*. General solution of D'Alembert's equation in the form of retarded potentials

We now pass on to the exact solution of D'Alembert's equation for the potentials. We confine ourselves to finding one of the potentials, for instance the vector potential. The expression for the scalar potential can be written by analogy.

In accordance with what was said in the preceding paragraph, the equation for the potential

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = - \frac{4\pi}{c} \mathbf{j} \quad (24.1)$$

must be supplemented by initial and boundary conditions.

We consider the problem in the following formulation. Before a certain time $t = 0$ (i.e. for $t < 0$) let there, for example, be a system of charges, at

rest or performing a steady state motion. Before the time $t = 0$ their configuration is given and the field has a certain distribution in space. At the moment $t = 0$ the state of the system is changed and the charges begin a non-steady state motion. We are interested in the law which gives the change of the field vectors, connected with this non-steady state motion of the charges, from the time $t = 0$ on.

Let \mathbf{E} and \mathbf{H} be field vectors associated with the motion of the charges for a time $t \geq 0$. In other words, let \mathbf{E} and \mathbf{H} be the changes in the electromagnetic field which existed at the moment $t = 0$, caused by the motion of the charges for $t \geq 0$. Then from the definition of \mathbf{E} and \mathbf{H} it follows that at the initial moment $t = 0$ they must satisfy the initial condition:

$$\mathbf{E} = \mathbf{H} = 0 \quad \text{at } t = 0 \text{ (over all space).}$$

In this case the initial condition for the vector potential has the following form:

$$\mathbf{A} = 0 \quad \text{at } t = 0, \quad \text{for arbitrary } r, \quad (24.2)$$

$$\frac{\partial \mathbf{A}}{\partial t} = 0 \quad \text{at } t = 0, \quad \text{for arbitrary } r. \quad (24.3)$$

The first of these equalities corresponds to the equality $\mathbf{H} = 0$, while the second one corresponds to the absence of the electric field, $\mathbf{E} = 0$, at the initial moment. The boundary condition at infinity has the form

$$\mathbf{A} \rightarrow O\left(\frac{1}{r}\right) \quad \text{as } r \rightarrow \infty, \quad t \geq 0. \quad (24.4)$$

In other words, we understand \mathbf{A} to be the vector potential of the field associated with the motion of charges beginning at the time $t = 0$ and taking place for $t > 0$.

To obtain the solution of this problem — i.e. finding the solution of eq. (24.1) satisfying the system of conditions (24.2)–(24.4) (called in mathematical physics the Cauchy problem) — it is most convenient to make use of the method of Fourier integrals*. We write the vector potential and current

* R. Courant and D. Hilbert, *Methods of mathematical physics* (Interscience Publishers, New York, 1953);

A. A. Vlasov, *Makroskopicheskaya elektrodinamika (Macroscopic electrodynamics)* (Gostekhizdat, Moscow, 1955) p. 166.

density in the form of the triple Fourier integrals *.

$$\mathbf{A} = \frac{1}{(2\pi)^{3/2}} \int \mathbf{a}(\mathbf{k}, t) e^{-i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k}, \quad (24.5)$$

$$\mathbf{j} = \frac{1}{(2\pi)^{3/2}} \int \mathbf{p}(\mathbf{k}, t) e^{-i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k}. \quad (24.6)$$

The Fourier transformations of the vector potential and current density are then given by the inversion formulae:

$$\mathbf{a}(\mathbf{k}, t) = \frac{1}{(2\pi)^{3/2}} \int \mathbf{A}(\mathbf{r}, t) e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{r}, \quad (24.7)$$

$$\mathbf{p}(\mathbf{k}, t) = \frac{1}{(2\pi)^{3/2}} \int \mathbf{j}(\mathbf{r}, t) e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{r}. \quad (24.8)$$

Substituting (24.5) and (24.6) into (24.1) we find

$$\frac{d^2 \mathbf{a}}{dt^2} + k^2 c^2 \mathbf{a} = 4\pi c \mathbf{p}. \quad (24.9)$$

Eq. (24.9) represents an equation in total derivatives, and its general solution can be found easily. Substitution of (24.7) into the initial conditions (24.2) and (24.3) gives:

$$\mathbf{a} = 0 \quad \text{at } t = 0, \quad (24.10)$$

$$\frac{d\mathbf{a}}{dt} = 0 \quad \text{at } t = 0. \quad (24.11)$$

The solution of eq. (24.9) with the initial conditions (24.10) and (24.11) reads **:

$$\mathbf{a} = \frac{4\pi}{k} \int_0^t \mathbf{p}(\xi) \sin [kc(t - \xi)] d\xi. \quad (24.12)$$

* In this section $d\mathbf{r}$ stands for the volume element $dx dy dz$ and $d\mathbf{k}$ for $dk_x dk_y dk_z$.

** V. I. Smirnov, *A course of higher mathematics* (Pergamon, Oxford, 1964).

To obtain the function A this value of a must be substituted into (24.5). We then have

$$A = \frac{4\pi}{(2\pi)^{3/2}} \int_0^t \int_0^\infty p(\xi) \sin \{kc(t-\xi)\} e^{-ik \cdot r} d\xi \frac{dk}{k}. \quad (24.13)$$

The current density j should be introduced instead of the Fourier transform p in (24.13). By means of (24.8) we find

$$A = \frac{4\pi}{(2\pi)^3} \int e^{-ik \cdot r} \frac{dk}{k} \int_0^t \int e^{ik \cdot r'} j(r', \xi) \sin \{kc(t-\xi)\} dr' d\xi. \quad (24.14)$$

Integration in formula (24.14) must be carried out with respect to the variables ξ and k .

We change the order of integration, writing

$$A = \frac{1}{2\pi^2} \int dr' \int_0^t j(r', \xi) d\xi \int e^{-ik \cdot (r-r')} \frac{\sin \{kc(t-\xi)\}}{k} dk. \quad (24.15)$$

We calculate the inner integral and have

$$\begin{aligned} & \int e^{-ik \cdot (r-r')} \frac{\sin \{kc(t-\xi)\}}{k} dk = \\ &= \int e^{-ik |r-r'| \cos \theta} \frac{\sin \{kc(t-\xi)\}}{k} k^2 dk \sin \theta d\theta d\varphi = \\ &= 2\pi \int_0^\infty \frac{\sin \{kc(t-\xi)\}}{k} k^2 dk \int_0^\pi e^{-ik |r-r'| \cos \theta} \sin \theta d\theta. \end{aligned}$$

But

$$\begin{aligned} \int_0^\pi e^{-ik |r-r'| \cos \theta} \sin \theta d\theta &= \int_{-1}^{+1} e^{-ik |r-r'| u} du = \\ &= 2 \frac{\sin \{k |r-r'| \}}{k |r-r'|} \end{aligned}$$

Hence

$$\begin{aligned} \int e^{-i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} \frac{\sin \{kc(t - \xi)\}}{k} d\mathbf{k} &= \\ &= \frac{4\pi}{|\mathbf{r} - \mathbf{r}'|} \int_0^\infty \sin \{kc(t - \xi)\} \sin \{k|\mathbf{r} - \mathbf{r}'|\} dk = \frac{4\pi}{|\mathbf{r} - \mathbf{r}'|} I, \end{aligned}$$

where I stands for the integral

$$\begin{aligned} I &= \int_0^\infty \sin \{kc(t - \xi)\} \sin \{k|\mathbf{r} - \mathbf{r}'|\} dk = \\ &= \int_0^\infty \sin \alpha k \sin \beta k dk = \frac{1}{2} \int_0^\infty \cos(\alpha - \beta) k dk - \\ &\quad - \frac{1}{2} \int_0^\infty \cos(\alpha + \beta) k dk = \frac{1}{2} \pi \delta(\alpha - \beta) - \frac{1}{2} \pi \delta(\alpha + \beta) = \\ &= \frac{1}{2} \pi \delta[c(t - \xi) - |\mathbf{r} - \mathbf{r}'|] - \frac{1}{2} \pi \delta[c(t - \xi) + |\mathbf{r} - \mathbf{r}'|]. \end{aligned}$$

In calculating I we have made use of one of the determinations of the delta function given in Appendix III.

Finally we find

$$\begin{aligned} \int e^{-i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} \frac{\sin \{kc(t - \xi)\}}{k} d\mathbf{k} &= \\ &= \frac{2\pi^2}{|\mathbf{r} - \mathbf{r}'|} \{ \delta[c(t - \xi) - |\mathbf{r} - \mathbf{r}'|] - \delta[c(t - \xi) + |\mathbf{r} - \mathbf{r}'|] \}. \end{aligned}$$

.Substituting this expression into (24.15), we obtain

$$\begin{aligned}
 A = & \int \frac{d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} \int_0^t \mathbf{j}(\mathbf{r}', \xi) \delta[c(t - \xi) - |\mathbf{r} - \mathbf{r}'|] d\xi + \\
 & + \int \frac{d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} \int_0^t \mathbf{j}(\mathbf{r}', \xi) \delta[c(t - \xi) + |\mathbf{r} - \mathbf{r}'|] d\xi.
 \end{aligned}$$

We seek to perform the integration with respect to the variable ξ , making use of the properties of the delta function.

In the first integral:

$$\begin{aligned}
 & \int_0^t \mathbf{j}(\mathbf{r}', \xi) \delta[c(t - \xi) - |\mathbf{r} - \mathbf{r}'|] d\xi = \\
 & = -\frac{1}{c} \int_{ct - |\mathbf{r} - \mathbf{r}'|}^{-|\mathbf{r} - \mathbf{r}'|} \mathbf{j}\left(\mathbf{r}', -\frac{u}{c} + t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) \delta(u) du = \\
 & = -\frac{1}{c} \int_{\infty}^{-\infty} \mathbf{j}\left(\mathbf{r}', -\frac{u}{c} + t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) \delta(u) du = \\
 & = \frac{1}{c} \mathbf{j}\left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right).
 \end{aligned}$$

We have chosen $u = c(t - \xi) - |\mathbf{r} - \mathbf{r}'|$ as a new variable.

If the inequality $t > |\mathbf{r} - \mathbf{r}'|/c$ holds, then the range of integration can be extended, as has been done here. If $t < |\mathbf{r} - \mathbf{r}'|/c$, then the integration would be carried out only with respect to negative values of the variable u (both limits would be negative). The point $u = 0$, at which $\delta(u)$ becomes infinite, would lie outside the range of integration, and by virtue of (III.3') the integral would be reduced to zero. This refers, in particular, to the time $t = 0$, which corresponds to the fulfillment of the initial condition (24.2). Similarly the second integral gives

$$\begin{aligned}
& \int_0^t \mathbf{j}(\mathbf{r}', \xi) \delta[c(t - \xi) + |\mathbf{r} - \mathbf{r}'|] d\xi = \\
& = -\frac{1}{c} \int_{ct + |\mathbf{r} - \mathbf{r}'|}^{|\mathbf{r} - \mathbf{r}'|} \mathbf{j}\left(\mathbf{r}', -\frac{u}{c} + t + \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) \delta(u) du = 0
\end{aligned}$$

where in this case the point $u = 0$ turns out to be outside the range of integration.

Thus we have obtained

$$\mathbf{A}(\mathbf{r}, t) = \frac{1}{c} \int \frac{\mathbf{j}\left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \quad (24.16)$$

i.e. the expression already known for the retarded potential.

The integral with the second delta function corresponded to the advanced potential. The advanced potential gave no contribution to the vector potential only because, in obtaining the solution (24.12) of eq. (24.9), we made use of the initial conditions (24.10) and (24.11), which determined the form of $\mathbf{a}(\mathbf{k}, t)$.

If we wrote the general solution of eq. (24.9) without giving initial conditions, then the final expression for the potential would involve a linear combination of the retarded and advanced potentials, as well as the general solution of the homogeneous equation.

Later we shall need an expression for the potential in the special case when the dependence of $\mathbf{j}(\mathbf{r}, t)$ on time is expressed by a simple harmonic law:

$$\mathbf{j}(\mathbf{r}, t) = \mathbf{j}_0(\mathbf{r}) e^{i\omega t} \quad (24.17)$$

Substitution into (24.16) gives

$$\mathbf{A}(\mathbf{r}, t) = \frac{1}{c} e^{i\omega t} \int \frac{\mathbf{j}_0(\mathbf{r}') e^{-i\frac{\omega}{c} |\mathbf{r} - \mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' = \mathbf{A}_0(\mathbf{r}) e^{i\omega t}. \quad (24.18)$$

As is to be expected, the vector potential depends on time in the same way as the current. The amplitude $\mathbf{A}_0(\mathbf{r})$ is obviously equal to

$$A_0(\mathbf{r}) = \frac{1}{c} \int \frac{j_0(\mathbf{r}') e^{-i \frac{\omega}{c} |\mathbf{r} - \mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \quad (24.19)$$

On the other hand, substituting (24.17) and (24.18) into (24.1), we have

$$\nabla^2 A_0 + \frac{\omega^2}{c^2} A_0 = -\frac{4\pi}{c} j_0. \quad (24.20)$$

Thus, formula (24.19) gives the solution of eq. (24.20) which we shall encounter later.

In conclusion it should be noted that the solution of Poisson's equation represents a particular case of the problem. Assuming in (24.20) and (24.19) that $\omega = 0$ we arrive at the relations:

$$\nabla^2 A_0 = -\frac{4\pi}{c} j_0, \quad (24.21)$$

$$A_0 = \frac{1}{c} \int \frac{j_0(\mathbf{r}') d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|}. \quad (24.22)$$

The expression (24.22) can also be obtained by a direct solution of (24.21) by Fourier integral expansion. Indeed, assuming that

$$A_0 = \frac{1}{(2\pi)^{3/2}} \int e^{-i\mathbf{k} \cdot \mathbf{r}} a_0(\mathbf{k}) d\mathbf{k}, \quad (24.23)$$

$$j_0 = \frac{1}{(2\pi)^{3/2}} \int e^{-i\mathbf{k} \cdot \mathbf{r}} p_0(\mathbf{k}) d\mathbf{k}, \quad (24.24)$$

and substituting (24.23) and (24.24) into (24.21), we have

$$a_0 = \frac{4\pi p_0}{ck^2}. \quad (24.25)$$

Consequently,

$$A_0 = \frac{1}{(2\pi)^{3/2}} \int e^{-i\mathbf{k} \cdot \mathbf{r}} \left(\frac{4\pi p_0}{ck^2} \right) d\mathbf{k}. \quad (24.26)$$

By means of the inversion formula we obtain from (24.24) .

$$\mathbf{p}_0 = \frac{1}{(2\pi)^{3/2}} \int e^{i\mathbf{k} \cdot \mathbf{r}} \mathbf{j}_0(\mathbf{r}) d\mathbf{r} ,$$

Substitution of this expression into (24.26) leads to the formula

$$\begin{aligned} \mathbf{A}_0 &= \frac{4\pi}{(2\pi)^3} \frac{1}{c} \int \mathbf{j}_0(\mathbf{r}') d\mathbf{r}' \int e^{-i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} \frac{d\mathbf{k}}{k^2} = \\ &= \frac{1}{2\pi^2} \frac{1}{c} \int \mathbf{j}_0(\mathbf{r}') d\mathbf{r}' \int_0^\infty \int_0^\pi \int_0^{2\pi} e^{-i\mathbf{k} \cdot |\mathbf{r} - \mathbf{r}'| \cos \theta} \sin \theta d\theta d\varphi dk = \\ &= \frac{2}{\pi c} \int \mathbf{j}_0(\mathbf{r}') d\mathbf{r}' \int_0^\infty \frac{\sin \{k|\mathbf{r} - \mathbf{r}'|\}}{k|\mathbf{r} - \mathbf{r}'|} dk = \frac{1}{c} \int \frac{\mathbf{j}_0(\mathbf{r}') d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} , \end{aligned}$$

where we have made use of the equality

$$\int_0^\infty \frac{\sin \{k|\mathbf{r} - \mathbf{r}'|\}}{k|\mathbf{r} - \mathbf{r}'|} dk = \frac{1}{|\mathbf{r} - \mathbf{r}'|} \int_0^\infty \frac{\sin u}{u} du = \frac{\pi}{2|\mathbf{r} - \mathbf{r}'|} .$$

§ 25*. The field of a point charge moving arbitrarily

The case of a single point charge moving arbitrarily is an important example of the application of the formula for retarded potentials. If the velocity of the charge is equal to $\mathbf{v}_0(t)$, then its coordinate is $\mathbf{r}_0 = \int \mathbf{v}_0 dt$. The charge density and current density can be written in the form

$$\rho = e \delta(\mathbf{r}' - \mathbf{r}_0) ,$$

$$\mathbf{j} = \rho \mathbf{v}_0 = e \delta(\mathbf{r}' - \mathbf{r}_0) \mathbf{v}_0(t) .$$

Then the following expressions are obtained for the retarded potentials:

$$\begin{aligned}
 A(\mathbf{r}, t) &= \frac{e}{c} \int \frac{\delta(\mathbf{r}' - \mathbf{r}_0(\tau)) v_0(\tau) dV'}{|\mathbf{r} - \mathbf{r}'|} = \\
 &= \frac{e}{c} \int \frac{\delta\left(\mathbf{r}' - \mathbf{r}_0\left(t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right)\right) v_0\left(t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) dV'}{|\mathbf{r} - \mathbf{r}'|}, \quad (25.1)
 \end{aligned}$$

$$\begin{aligned}
 \varphi(\mathbf{r}, t) &= e \int \frac{\delta(\mathbf{r}' - \mathbf{r}_0(\tau)) dV'}{|\mathbf{r} - \mathbf{r}'|} = \\
 &= e \int \frac{\delta\left(\mathbf{r}' - \mathbf{r}_0\left(t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right)\right) dV'}{|\mathbf{r} - \mathbf{r}'|}. \quad (25.2)
 \end{aligned}$$

Since \mathbf{r}_0 in the integrand of formulae (25.1) and (25.2) is a function of the delay time $t - |\mathbf{r} - \mathbf{r}'|/c$, one cannot make use of the property of the delta-function directly and assume $\mathbf{r}' = \mathbf{r}_0$. In order to carry out the integration we introduce new variables:

$$l_x = x' - x_0(\tau), \quad l_y = y' - y_0(\tau), \quad l_z = z' - z_0(\tau).$$

The calculation of the Jacobian for the transition to the new variables is in general rather cumbersome. For simplicity we assume that the charge is moving along the axis x_0 , so that $v_x = v_0$. We then have

$$\begin{aligned}
 \frac{\partial l_x}{\partial x'} &= 1 - \frac{\partial x_0}{\partial x'} = 1 - \frac{\partial x_0}{\partial \tau} \frac{\partial \tau}{\partial x'} = \\
 &= 1 - v_0 \frac{\partial}{\partial x'} \left\{ t - \frac{|\mathbf{r} - \mathbf{r}'|}{c} \right\} = 1 - \frac{v_0}{c} \frac{x - x'}{|\mathbf{r} - \mathbf{r}'|};
 \end{aligned}$$

$$\frac{\partial l_x}{\partial y'} = 0; \quad \frac{\partial l_x}{\partial z'} = 0;$$

$$\frac{\partial l_y}{\partial x'} = 0; \quad \frac{\partial l_y}{\partial y'} = 1; \quad \frac{\partial l_y}{\partial z'} = 0;$$

$$\frac{\partial l_z}{\partial x'} = 0; \quad \frac{\partial l_z}{\partial y'} = 0; \quad \frac{\partial l_z}{\partial z'} = 1.$$

Thus, the Jacobian is equal to

$$\frac{\partial(l_x, l_y, l_z)}{\partial(x', y', z')} = 1 - \frac{v_0(x - x')}{c|\mathbf{r} - \mathbf{r}'|}$$

An analogous calculation for an arbitrary orientation of the velocity leads to the expression

$$\frac{\partial(l_x, l_y, l_z)}{\partial(x', y', z')} = 1 - \frac{\mathbf{v}_0 \cdot (\mathbf{r} - \mathbf{r}')}{c|\mathbf{r} - \mathbf{r}'|}.$$

Hence one can write

$$\begin{aligned} dx' dy' dz' &= \frac{\partial(x', y', z')}{\partial(l_x, l_y, l_z)} dl_x dl_y dl_z = \\ &= \frac{dl_x dl_y dl_z}{1 - \frac{\mathbf{v}_0 \cdot (\mathbf{r} - \mathbf{r}')}{c|\mathbf{r} - \mathbf{r}'|}} = \frac{dl}{1 - \frac{\mathbf{v}_0 \cdot (\mathbf{r} - \mathbf{r}')}{c|\mathbf{r} - \mathbf{r}'|}}. \end{aligned}$$

Transforming the expressions (25.1) and (25.2) to new variables, we obtain

$$\begin{aligned} A(\mathbf{r}, t) &= \frac{e}{c} \int \frac{\delta(l) v_0 \left[t - \frac{|\mathbf{r} - \mathbf{l} - \mathbf{r}_0|}{c} \right] dl}{|\mathbf{r} - \mathbf{l} - \mathbf{r}_0| - \frac{(\mathbf{r} - \mathbf{l} - \mathbf{r}_0) \cdot \mathbf{v}_0}{c} \left[t - \frac{|\mathbf{r} - \mathbf{l} - \mathbf{r}_0|}{c} \right]} = \\ &= \frac{e v_0}{c \left[|\mathbf{r} - \mathbf{r}_0| - \frac{\mathbf{v}_0 \cdot (\mathbf{r} - \mathbf{r}_0)}{c} \right]} = \frac{e v_0(\tau)}{c \left[R(\tau) - \frac{\mathbf{v}_0(\tau) \cdot \mathbf{R}(\tau)}{c} \right]}, \end{aligned} \quad (25.3)$$

where $\mathbf{R}(\tau)$ stands for the radius-vector drawn from the instantaneous position of the charge to the point of observation, i.e. $\mathbf{R}(\tau) = \mathbf{r} - \mathbf{r}_0$. The value of the instantaneous position of the charge must be taken at the time τ :

$$\tau = t - \frac{|\mathbf{r} - \mathbf{r}_0|}{c}$$

The value of the instantaneous velocity $\mathbf{v}_0(\tau)$ is also taken at the time τ . Analogously,

$$\begin{aligned}
 \varphi(\mathbf{r}, t) &= e \int \frac{\delta(l) dl}{|\mathbf{r} - \mathbf{l} - \mathbf{r}_0| - \frac{(\mathbf{r} - \mathbf{l} - \mathbf{r}_0) \cdot \mathbf{v}_0}{c} \cdot v_0 \left(t - \frac{|\mathbf{r} - \mathbf{l} - \mathbf{r}_0|}{c} \right)} = \\
 &= \frac{e}{|\mathbf{r} - \mathbf{r}_0| - \frac{\mathbf{v}_0(\tau) \cdot (\mathbf{r} - \mathbf{r}_0)}{c}} = \frac{e}{R(\tau) - \frac{\mathbf{v}_0(\tau) \cdot \mathbf{R}(\tau)}{c}}. \quad (25.4)
 \end{aligned}$$

Between φ and \mathbf{A} there is the relation

$$\mathbf{A} = \mathbf{v}_0 \varphi / c. \quad (25.5)$$

The potentials of the field of an arbitrarily moving point charge (25.3) and (25.4) are called the Liénard-Wiechert potentials. If the abbreviated notation

$$\lambda(\tau) = R(\tau) - \frac{\mathbf{v}_0(\tau) \cdot \mathbf{R}(\tau)}{c}, \quad (25.6)$$

is introduced, then the Liénard-Wiechert potentials assume the form

$$\mathbf{A} = e \mathbf{v}_0 / c \lambda, \quad (25.7)$$

$$\varphi = e / \lambda. \quad (25.8)$$

The importance of the Lienard-Wiechert potentials lies in the fact that they characterize the field of a point charge in the most general form — for an arbitrary value of the velocity and the character of the motion.

It is easy to see that for a velocity of a charge $|\mathbf{v}_0|$ which is very small in comparison with the velocity of light, i.e. for $|\mathbf{v}_0|/c \rightarrow 0$ the expression for the scalar potential φ goes over into formula (20.2), and the vector potential \mathbf{A} turns out to be small. However, in deriving the Liénard-Wiechert potentials we have put no restrictions upon the value of the velocity. Hence the Liénard-Wiechert potentials characterize the field of a point charge in the most general case — for an arbitrary character and velocity of motion.

The field of a charge moving arbitrarily will be discussed in more detail in § 20 of Part II, since a number of inferences from formulae (25.7)–(25.8) can become clear only in the light of the theory of relativity.

Radiation Theory

§ 26. The potentials of the electromagnetic field at a large distance from the emitter in the dipole approximation

The general formulae for retarded potentials obtained in § 23 are very complex. Indeed, since the expressions for the charge density and current density contained in (23.22) and (23.23) are functions of the delay time in the corresponding integrals, it is necessary to take the values of these quantities at different times at each point of the system for the calculation of potentials at time t . Hence, except for the case of a single point charge considered in the preceding paragraph, one is unable to obtain accurate concrete expressions for the potentials by means of the general formulae (23.22) and (23.23).

If, however, the point of observation is located at a sufficiently large distance from the system of moving charges, so that $|\mathbf{r}| \gg L'$, where $L' \sim |V'|^{1/3}$ is the characteristic linear dimension of the system, then the expressions (23.22) and (23.23) allow a simplification. Namely, the expression $1/|\mathbf{r} - \mathbf{r}'|$ figuring in the integrals can be expanded in a series, as was done in calculating the fields of motionless (§ 15) and slowly moving (§ 21) charges. We then obtain

$$\begin{aligned}
 \varphi &= \int \frac{\rho\left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) dV'}{|\mathbf{r} - \mathbf{r}'|} \approx \int \left(\frac{1}{r} + \frac{\mathbf{r}' \cdot \mathbf{r}}{r^3}\right) \rho\left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) dV' = \\
 &= \int \frac{\rho\left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) dV'}{r} \\
 &\quad + \int \frac{\mathbf{r}' \cdot \mathbf{r}}{r^3} \rho\left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) dV', \quad (26.1)
 \end{aligned}$$

where r is the distance from the point of observation N to the origin O.

It should be noted that $\int \rho(\mathbf{r}', \tau) dV'$ is by no means the total charge of the system, i.e.

$$e \neq \int \rho(\mathbf{r}', \tau) dV'.$$

Indeed, the value of the density in this integral depends on the argument $\tau = t - |\mathbf{r} - \mathbf{r}'|/c$ and represents a complicated function of coordinates \mathbf{r} , \mathbf{r}' and time. The delay time is different for each point in a volume V' . For this reason the integrals in (26.1) cannot be calculated in a general form.

A further simplification arises in the case where instead of a different delay time for each point of the system one can introduce one general delay time for the entire system. Namely, writing the argument $\tau = t - |\mathbf{r} - \mathbf{r}'|/c$ in the form

$$\tau \approx t - \frac{r}{c} + \frac{\mathbf{r}' \cdot \mathbf{r}}{cr} = \tau_0 + \frac{\mathbf{r}' \cdot \mathbf{r}}{cr}, \quad (26.2)$$

we see that the total delay time $|\mathbf{r} - \mathbf{r}'|/c$ is made up of two parts. The first one, equal to r/c and called the delay time of the system, represents the time needed for the propagation of the electromagnetic field from the origin O to the point of observation N. The second part, equal to $\mathbf{r}' \cdot \mathbf{r}/cr$ and called the proper retardation, also has a simple meaning: it is the time needed for the propagation of the field to the limits of the system. According to the order of magnitude $\mathbf{r}' \cdot \mathbf{r}/cr \sim L'/c$, and for $|\mathbf{r}| \gg L'$ the absolute value of the proper retardation $|\mathbf{r}'|/c$ is small in comparison with $|\mathbf{r}|/c$. However, this does not mean that the charge density can be expanded in a series in the small parameter $\mathbf{r}' \cdot \mathbf{r}/cr$.

$$\begin{aligned}
 \rho(\mathbf{r}', \tau) &= \rho\left(\mathbf{r}', t - \frac{|\mathbf{r} - \mathbf{r}'|}{c}\right) = \\
 &= \rho\left(\mathbf{r}', t - \frac{r}{c} + \frac{\mathbf{r}' \cdot \mathbf{r}}{cr}\right) \approx \rho\left(\mathbf{r}', t - \frac{r}{c}\right) + \frac{\partial \rho}{\partial \tau} \frac{\mathbf{r}' \cdot \mathbf{r}}{cr} = \\
 &= \rho(\mathbf{r}', \tau_0) + \frac{\mathbf{r}' \cdot \mathbf{r}}{cr} \dot{\rho}(\mathbf{r}', \tau_0). \quad (26.3)
 \end{aligned}$$

Indeed, if in a time equal to the proper delay time $\mathbf{r}' \cdot \mathbf{r}/cr$ the configuration of the charges in the system manages to change appreciably, i.e. if in this time the charges manage to be displaced appreciably in the system, then the charge density at a moment $t - r/c$ will differ substantially from the charge density at a moment

$$t - \frac{r}{c} + \frac{\mathbf{r}' \cdot \mathbf{r}}{cr}.$$

In other words, the charge density ρ will be a rapidly varying function of its argument, and it is inadmissible to make use of the equality (26.3). In order that this equality may hold, it is necessary that in the time $\mathbf{r}' \cdot \mathbf{r}/cr$, in the course of which a field propagating with velocity c crosses the system, the charges in the system moving with a velocity v will not be displaced appreciably. During the time $\mathbf{r}' \cdot \mathbf{r}/cr$ the charges traverse a path of the order of

$$v \frac{\mathbf{r}' \cdot \mathbf{r}}{cr} \sim v \frac{L'}{c}.$$

If this path is small in comparison with the dimensions of the system, it can be assumed that during the time of proper retardation the disposition of the charges in the system does not change.

Thus, it can be assumed that for

$$v(L'/c) \ll L',$$

or at the velocities of motion satisfying the inequality

$$v \ll c, \quad (26.4)$$

the change of the configuration in the time of proper retardation is small.

Then $\rho(\mathbf{r}', \tau)$ is a slowly varying function of its argument. This means that to small changes in τ there correspond small changes in ρ , and use can be made of the expansion of ρ in powers of the small retardation. Substituting (26.3) into (26.1) and confining ourselves to the terms of the expansion containing the lowest powers of $1/r$, we find

$$\begin{aligned}\varphi &\approx \int \left(\frac{1}{r} + \frac{\mathbf{r}' \cdot \mathbf{r}}{r^3} \right) \left(\rho(\mathbf{r}', \tau_0) + \frac{\mathbf{r}' \cdot \mathbf{r}}{cr} \dot{\rho}(\mathbf{r}', \tau_0) \right) dV' \approx \\ &\approx \int \left\{ \frac{\rho(\mathbf{r}', \tau_0)}{r} + \frac{\mathbf{r}' \cdot \mathbf{r}}{cr^2} \dot{\rho}(\mathbf{r}', \tau_0) \right\} dV' = \\ &= \int \frac{\rho(\mathbf{r}', \tau_0) dV'}{r} + \frac{\mathbf{n}}{cr} \cdot \int \mathbf{r}' \dot{\rho}(\mathbf{r}', \tau_0) dV',\end{aligned}\quad (26.5)$$

where $\mathbf{n} = \mathbf{r}/r$. The term

$$\frac{\mathbf{r}' \cdot \mathbf{r}}{r^3} \rho(\mathbf{r}', \tau_0)$$

is small in comparison with the term

$$\frac{\mathbf{r}' \cdot \mathbf{r}}{cr^2} \dot{\rho}(\mathbf{r}', \tau_0)$$

at a sufficiently large distance from the system.

In formula (26.5) a very substantial simplification is made in comparison with (26.1), since the charge density at all points of the system is taken at one and the same moment

$$\tau_0 = t - (r/c).$$

The first term in (26.5) has a simple meaning:

$$\rho(\mathbf{r}', \tau_0) = \rho\left(\mathbf{r}', t - \frac{r}{c}\right)$$

represents the charge density in the system at the moment τ_0 . The integral $\int \rho(\mathbf{r}', \tau_0) dV'$ gives the total charge of the system. For an electrically neutral system it is equal to zero. In this case we have

$$\varphi(\mathbf{r}, t) = \frac{n}{cr} \cdot \int \mathbf{r}' \cdot \dot{\rho}(\mathbf{r}', \tau_0) dV' . \quad (26.6)$$

We rewrite the integral on the right-hand side of formula (26.6) by making use of the continuity equation (5.3):

$$\int \mathbf{r}' \cdot \dot{\rho}(\mathbf{r}', \tau_0) dV' = \int \mathbf{r}' \cdot \frac{\partial \rho}{\partial t} \frac{\partial t}{\partial \tau_0} dV' = - \int \mathbf{r}' \cdot [\nabla \cdot \mathbf{j}(\mathbf{r}', \tau_0)] dV' .$$

The above integral is conveniently calculated in the coordinate representation. Namely, we have

$$\int x' \frac{\partial j_{x'}}{\partial x'} dx' = x' j_{x'} \Big|_{x'_1}^{x'_2} - \int j_{x'} dx' ,$$

where x'_1 and x'_2 are the limits of the region of motion of the charges, at which the current density reduces to zero, so that

$$\int x' \frac{\partial j_{x'}}{\partial x'} dx' = - \int j_{x'} dx' .$$

Correspondingly, in the vector form we obtain

$$\int \mathbf{r}' \cdot \dot{\rho}(\mathbf{r}', \tau_0) dV' = \int \mathbf{j} dV' . \quad (26.7)$$

Substituting (26.7) into (26.6), we find the scalar potential as a function of the current density in the system

$$\varphi(\mathbf{r}, t) = \frac{n}{cr} \cdot \int \mathbf{j}(\mathbf{r}', \tau_0) dV' . \quad (26.8)$$

In an analogous way we can obtain the expression for the vector-potential by expanding the expression $1/|\mathbf{r}-\mathbf{r}'|$ in formula (23.23) in a series and disregarding the proper retardation:

$$\mathbf{A}(\mathbf{r}, t) = \frac{1}{cr} \int \mathbf{j}(\mathbf{r}', \tau_0) dV' . \quad (26.9)$$

Comparing (26.8) and (26.9) we find that there is between φ and \mathbf{A} the simple relation:

$$\varphi = \mathbf{A} \cdot \mathbf{n}. \quad (26.10)$$

The integral $\int \mathbf{r}' \dot{\rho}(\mathbf{r}', \tau_0) dV'$ has a simple meaning. Namely, from the definition of the dipole moment (15.14) we see that

$$\int \mathbf{j} dV' = \int \mathbf{r}' \dot{\rho}(\mathbf{r}', \tau_0) dV' = \frac{\partial}{\partial \tau} \int \mathbf{r}' \rho(\mathbf{r}', \tau_0) dV' = \dot{\mathbf{d}}(\tau_0), \quad (26.11)$$

where $\dot{\mathbf{d}}(\tau_0)$ is the derivative of the dipole moment with respect to time, taken at the instant τ_0 . Here we have used the fact that \mathbf{r}' is an independent integration variable which does not depend on τ_0 .

By means of (26.11) the expressions (26.8) and (26.9) can be written in the form

$$\varphi(\mathbf{r}, t) = \frac{\mathbf{n} \cdot \dot{\mathbf{d}}(\tau_0)}{cr} \quad (26.12)$$

$$\mathbf{A}(\mathbf{r}, t) = \frac{\dot{\mathbf{d}}(\tau_0)}{cr} \quad (26.13)$$

We see that in the approximation when the proper retardation can be disregarded the field potentials at a large distance from the system are determined by the value of the derivative of its dipole moment with respect to time. Hence such an approximation in calculating field potentials is called the dipole approximation. The condition of applicability of the dipole approximation is the fulfillment of the inequality (26.4).

In the dipole approximation field potentials at a large distance from an electrically neutral system decrease according to the law $1/r$, whereas the analogous electrostatic potential of an electrically neutral system of charges at rest possessing a dipole moment varies according to the law $1/r^2$.

It is easy to verify that the Lorentz condition is fulfilled for the potentials found in the dipole approximation. We have

$$\begin{aligned} \nabla \cdot \mathbf{A} &= \frac{1}{c} \nabla \cdot \frac{\dot{\mathbf{d}}(\tau_0)}{r} = \\ &= \frac{1}{cr} \nabla \cdot \dot{\mathbf{d}}(\tau_0) + \frac{1}{c} \left(\dot{\mathbf{d}}(\tau_0) \cdot \nabla \frac{1}{r} \right) \approx \frac{1}{cr} \nabla \cdot \dot{\mathbf{d}}(\tau_0). \end{aligned} \quad (26.14)$$

In calculating (26.14), the second term in the sum, which is proportional to

$1/r^2$ is small in comparison with the first which is proportional to $1/r$, and can be discarded. We see that in differentiating with respect to coordinates at a large distance from the emitter the quantity $1/r$ can be considered to be constant. Further, according to formula (I.39) we have

$$\nabla \cdot \dot{\mathbf{d}}(\tau_0) = \frac{\partial \dot{\mathbf{d}}(\tau_0)}{\partial \tau_0} \cdot \nabla \tau_0 = -\ddot{\mathbf{d}} \cdot \frac{\mathbf{r}}{cr} = -\frac{\ddot{\mathbf{d}} \cdot \mathbf{n}}{c}.$$

Thus

$$\nabla \cdot \mathbf{A} = -\frac{1}{c^2 r} \ddot{\mathbf{d}}(\tau_0) \cdot \mathbf{n}.$$

On the other hand,

$$\frac{1}{c} \frac{\partial \varphi}{\partial t} = \frac{1}{c^2} \frac{\partial}{\partial t} \frac{\dot{\mathbf{d}} \cdot \mathbf{n}}{r} \approx \frac{1}{c^2} \frac{\mathbf{n}}{r} \cdot \ddot{\mathbf{d}}(\tau_0),$$

hence

$$\nabla \cdot \mathbf{A} + \frac{1}{c} \frac{\partial \varphi}{\partial t} = 0,$$

as was to be expected.

The results obtained have a simple and very important meaning: when charges in a system are moving (i.e. when the dipole moment of the system changes in time) an electromagnetic field arises in the surrounding space. The potentials of this field decrease relatively slowly (according to the law $1/r$) with increasing distance from the system and depend on time.

A system of non-uniformly moving charges is an emitter of radiation.

In the following paragraphs we shall consider in more detail the radiation field and the properties of systems emitting radiation.

§27. The electromagnetic field of dipole radiation at a large distance from the emitter

Knowing the potential distribution, one can find the values of the magnetic and electric field. We have

$$\mathbf{H} = \nabla \times \mathbf{A} = \frac{1}{c} \nabla \times \frac{\dot{\mathbf{d}}(\tau_0)}{r}.$$

In calculating the curl at a large distance from the emitter the calculation is to be carried out in the same way as for the divergence in formula (26.14): in differentiating with respect to coordinates the factor $1/r$ should be assumed to be constant. Then, according to formula (1.40), we find

$$\mathbf{H} = \frac{1}{cr} \nabla \times \dot{\mathbf{d}}(\tau_0) = \frac{1}{cr} \nabla \tau_0 \times \frac{d\dot{\mathbf{d}}(\tau_0)}{d\tau_0} = \frac{1}{c} \dot{\mathbf{A}} \times \mathbf{n} = \frac{1}{c^2 r} \ddot{\mathbf{d}} \times \mathbf{n}. \quad (27.1)$$

We have omitted writing \mathbf{d} explicitly as a function of τ_0 ; however, here and in all subsequent relations of this section we understand that \mathbf{d} is a function of the argument $\tau_0 = t - r/c$.

For any function of the retarded argument $t - r/c$ we have

$$\nabla f\left(t - \frac{r}{c}\right) = \frac{df}{d\tau_0} \nabla \tau_0 = -\dot{f} \frac{\mathbf{n}}{c},$$

where it is assumed that

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial \tau_0} \frac{\partial \tau_0}{\partial t} = \dot{f}.$$

Hence for the electric field

$$\begin{aligned} \mathbf{E} &= -\frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} - \nabla \varphi = -\frac{1}{c} \dot{\mathbf{A}} + \frac{\mathbf{n}}{c} \dot{\varphi} = -\frac{1}{c} \dot{\mathbf{A}} + \frac{\mathbf{n}}{c} (\mathbf{n} \cdot \dot{\mathbf{A}}) = \\ &= \frac{1}{c} \{ \mathbf{n}(\mathbf{n} \cdot \dot{\mathbf{A}}) - \dot{\mathbf{A}} \} = \frac{1}{c} (\dot{\mathbf{A}} \times \mathbf{n}) \times \mathbf{n} = \frac{1}{c^2 r} (\ddot{\mathbf{d}} \times \mathbf{n}) \times \mathbf{n}. \end{aligned} \quad (27.2)$$

Comparing (27.2) with (27.1) we see that the vectors \mathbf{E} and \mathbf{H} are connected by the relation

$$\mathbf{E} = \mathbf{H} \times \mathbf{n}. \quad (27.3)$$

The strengths of the electric and magnetic field depend on coordinates and time according to the law

$$|\mathbf{H}| = |\mathbf{E}| = \frac{f(t - r/c)}{r}.$$

As we have seen in §23, the above formula is an expression for a spherical wave. The amplitude of the wave decreases at a large distance from the emitter according to the law $\sim 1/r$. In this case the vectors of the electric and magnetic field are equal to each other in absolute value, and are perpendicular to each other and to the radius vector \mathbf{r} .

The region at a large distance from the emitter, in which the electromagnetic field is described by spherical waves, is called the radiation zone. Somewhat later we shall define more precisely the notion of the radiation zone.

We introduce the spherical system of coordinates r, θ, ψ (fig. I.10) with the polar axis directed along the vector \mathbf{d} . The direction of the vector \mathbf{H} is determined by the vector $[\mathbf{n} \times \mathbf{d}]$, directed along the tangent to the latitude line on the surface of the sphere and oriented in the direction of decreasing azimuthal angle ψ , so that

$$[\mathbf{n} \times \ddot{\mathbf{d}}]_r = 0, \quad [\mathbf{n} \times \ddot{\mathbf{d}}]_\theta = 0, \quad [\mathbf{n} \times \ddot{\mathbf{d}}]_\psi = -\ddot{d} \sin \theta.$$

Hence the vector \mathbf{H} in spherical coordinates has the following components:

$$H_r = 0, \quad H_\theta = 0, \quad H_\psi = \frac{\ddot{d}}{c^2 r} \sin \theta. \quad (27.4)$$

The vector \mathbf{E} is directed perpendicularly to the vectors \mathbf{H} and \mathbf{n} along the tangent to the line of longitude, and is oriented in the direction of decreasing

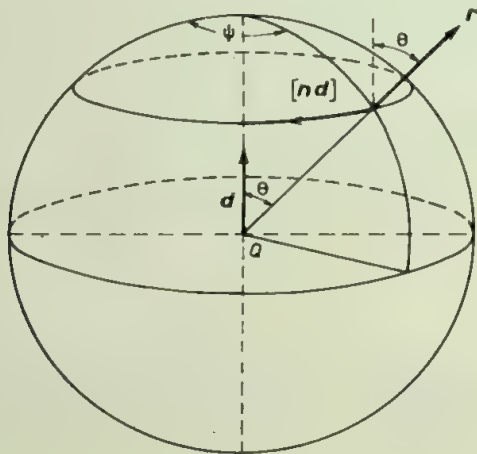


Fig. I.10



polar angle θ . Its components are

$$E_r = 0, \quad E_\theta = \frac{\ddot{d}}{c^2 r} \sin \theta, \quad E_\psi = 0. \quad (27.5)$$

Formulae (27.4) and (27.5) show that the field strengths have the largest values at $\theta = \pi/2$ (in the equatorial plane) and decrease down to zero as the distance from the polar axis decreases.

We calculate the Poynting vector of the emitting system:

$$\begin{aligned} \sigma &= \frac{c}{4\pi} \mathbf{E} \times \mathbf{H} = \frac{c}{4\pi} H^2 \mathbf{n} = \frac{c}{8\pi} (E^2 + H^2) \mathbf{n} = \\ &= cu_0 \mathbf{n} = \frac{1}{4\pi c} [\dot{\mathbf{A}} \times \mathbf{n}]^2 \mathbf{n} = \frac{1}{4\pi c^3 r^2} [\mathbf{n} \times \dot{\mathbf{d}}]^2 \mathbf{n}. \end{aligned} \quad (27.6)$$

The Poynting vector turns out to be directed along the radius-vector and is equal in absolute value to

$$\sigma = \frac{1}{4\pi c^3} \frac{\ddot{d}^2}{r^2} \sin^2 \theta. \quad (27.7)$$

The fact that the Poynting vector is different from zero and is always directed away from the emitting system has an obvious meaning: there is a flux of electromagnetic energy emitted by the system into the surrounding space. Formula (27.7) determines the flux density of the emitted energy as a function of the orientation in space (the angle θ) and the distance from the emitting system. The presence of the energy flux justifies the terms "radiation" and "emitter", which we introduced earlier.

We stress that in a factual observation of radiation in a certain direction \mathbf{n} , only the value of the component of the vector of the second derivative of the dipole moment ($\ddot{\mathbf{d}}$) in the plane perpendicular to \mathbf{n} is effective, as is seen from (27.6).

The energy flux through the vector surface $d\Sigma$ enclosing a solid angle $d\Omega$ is the power radiated in the solid angle $d\Omega$. For the power radiated dI

$$\begin{aligned} dI &= \sigma \cdot d\Sigma = \sigma d\Sigma = \sigma r^2 d\Omega = \frac{c}{4\pi} H^2 r^2 d\Omega = \frac{\ddot{d}^2 \sin^2 \theta}{4\pi c^3 r^2} r^2 d\Omega = \\ &= \frac{|\ddot{\mathbf{d}}|^2}{4\pi c^3} \sin^3 \theta d\theta d\psi = \frac{[\mathbf{n} \times \ddot{\mathbf{d}}]^2}{4\pi c^3} d\Omega. \end{aligned} \quad (27.8)$$

The total flux of energy emitted by the system, usually called the total radiated power, is equal to

$$-\frac{dE}{dt} = I = \int \sigma \cdot d\Sigma = \frac{\ddot{d}^2}{4\pi c^3} \int_0^\pi \sin^3 \theta \, d\theta \int_0^{2\pi} d\psi = \frac{2}{3} \frac{\ddot{d}^2}{c^3} \quad (27.9)$$

Here $(-dE/dt)$ is the decrease in the energy of the emitting system per second.

The power radiated in the dipole approximation is determined only by the value of $\ddot{d}(t-r/c)$. In other words, at a time t the value of I at a given point depends on the value of \ddot{d} at the preceding time $t-r/c$. But in other respects the power radiated does not depend on the distance from the emitting system, as is to be expected on the basis of the energy conservation law: the energy flux passing per unit time through any closed surface surrounding the emitting system has one and the same value.

In conclusion we point to the relation holding for the density of the energy and momentum of the radiation. By virtue of formulae (27.6) and (13.11) we can write

$$\mathbf{g} = \frac{1}{c^2} \boldsymbol{\sigma} = \frac{u_0}{c} \mathbf{n}. \quad (27.10)$$

We shall often encounter this important relation later.

We see that in emitting radiation the emitting system loses both energy and momentum, which become the energy and momentum of the radiation.

§ 28. Dipole radiation of simple systems

As an example of the use of dipole radiation formulae we shall consider several simple systems.

For a single charge, accelerated by a force \mathbf{F} , we can write

$$\left. \begin{aligned} \mathbf{d} &= e\mathbf{r}, \\ \ddot{\mathbf{d}} &= e\ddot{\mathbf{r}} = e \frac{\mathbf{F}}{m}, \end{aligned} \right\} \quad (28.1)$$

where $\mathbf{r} = \mathbf{r}$ is the acceleration of the charge. Hence

$$\sigma = \frac{c}{4\pi} H^2 \mathbf{n} = \frac{e^2 w^2 \sin^2 \theta}{4\pi c^3 r^2} \mathbf{n}. \quad (28.2)$$

Energy flux is absent in the direction of the acceleration vector ($\theta = 0$) and has its largest value in the direction perpendicular to the acceleration vector ($\theta = \pi/2$). The total energy emitted per unit time in the solid angle $d\Omega$ (total power of radiation in the angle $d\Omega$) is equal to

$$dI = \frac{e^2 w^2}{4\pi c^3} \sin^2 \theta d\Omega \quad (28.3)$$

and is proportional to the square of the acceleration. The total radiated power emitted by a single charge in all directions is equal to (see (27.9))

$$I = \frac{2}{3} \frac{e^2 w^2}{c^3}. \quad (28.4)$$

It is determined by the value of the square of the acceleration and the square of the charge.

We see that every charge moving with an acceleration emits energy in the form of electromagnetic waves.

The application of (28.2) to a single charge needs some explanation. In deriving formula (27.6) the expansion was carried out with respect to a small parameter — the proper time of retardation — which loses its meaning in the case of a system consisting of one charge.

However, in §25 we found the expressions for the field potentials of an arbitrarily moving point charge. If one substitutes in (25.7)

$$e\mathbf{v}_0 = \int \rho \mathbf{v}_0 dV = \int \mathbf{j} dV = \dot{\mathbf{d}},$$

and considers the case of motion with a velocity which is small in comparison with the velocity of light, so that $\lambda(\tau) \approx r$, then the expression for the vector-potential of the point charge is the same as formula (26.13).

Thus, at a large distance from a slowly moving charge its field will be the same as that of a system of charges in the dipole approximation.

Let us also calculate the loss of momentum of the emitting particle. By virtue of formulae (13.11) and (27.10) we see that the rate of loss of momentum of the charge while emitting is

$$-\frac{d\mathbf{P}}{dt} = \int \mathbf{g} d\Omega = \frac{1}{c^2} \int \sigma d\Omega = 0. \quad (28.5)$$

The meaning of this result lies in the fact that the emission by the charge at angles θ , $\pi - \theta$ is the same. The momenta in opposite directions are mutually compensated. It should be stressed that formulae (28.4) and (28.5) refer to a charge which was at rest at the moment of emission. It is only for such a charge that the relation (28.1) holds.

Let us consider several simple examples of the calculation of radiation from a moving point charge.

Let, for instance, the charge move in a uniform magnetic field. For simplicity we assume that the initial velocity of the charge \mathbf{v}_0 is perpendicular to the vector \mathbf{H} . The charge moving in the magnetic field has the acceleration

$$\mathbf{w} = \frac{\mathbf{F}}{m} = \frac{e}{mc} \mathbf{v} \times \mathbf{H}$$

and, correspondingly, continually emits electromagnetic waves. The total intensity of the radiation is equal to

$$I = \frac{2}{3} \frac{e^4}{m^2 c^5} [\mathbf{v} \times \mathbf{H}]^2 \quad (28.6)$$

If the energy loss is assumed to be small, then the velocity of the charge can be considered as approximately constant: $\mathbf{v} \approx \mathbf{v}_0$. In this case $\mathbf{v} \times \mathbf{H} \approx \mathbf{v}_0 \times \mathbf{H} = v_0 \mathbf{H}$. Hence

$$I = \frac{2}{3} \frac{e^4 v^2 H^2}{m^2 c^5} = \frac{4}{3} \frac{e^4 H^2}{m^3 c^5} \left(\frac{mv^2}{2} \right). \quad (28.6')$$

The energy emitted is inversely proportional to c^5 and is very small. However, it increases with charged particle energy and the effect of emission becomes substantial at very high particle energies, for example, for cosmic-ray particles in the Earth's magnetic field, or for fast electrons moving in the magnetic fields of contemporary betatrons. Calculations have shown that it is the energy losses due to radiation in a magnetic field which are the main losses limiting the energies attainable by particles in a betatron. However, it should be borne in mind that formula (28.6) is applicable only at velocities $v \ll c$. The case $v \sim c$ is discussed in §26 of Part II.

As another example, let us consider the radiation by a charge oscillating according to the harmonic law:

$$\mathbf{r} = \mathbf{r}_0 \cos(\omega_0 t + \alpha). \quad (28.7)$$

The acceleration of the charge is equal to

$$\ddot{\mathbf{r}} = -\omega_0^2 \mathbf{r} = -\omega_0^2 \mathbf{r}_0 \cos(\omega t + \alpha), \quad (28.8)$$

so that the corresponding dipole moment is expressed by $\mathbf{d} = \mathbf{d}_0(\cos \omega_0 t + \alpha)$. Hence the radiation power emitted in the solid angle $d\Omega$ is equal to

$$dI = \frac{e^2 \omega_0^4}{4\pi c^3} [\mathbf{n} \times \mathbf{r}_0]^2 \cos^2(\omega t + \alpha) d\Omega.$$

The mean (over one period) intensity of radiation in the angle $d\Omega$ is

$$d\bar{I} = \frac{1}{T} \int_0^T dI = \frac{e^2 \omega_0^4}{8\pi c^3} [\mathbf{n} \times \mathbf{r}_0]^2 d\Omega. \quad (28.9)$$

Formula (28.9) determines, in particular, the angular distribution of the radiation emitted. The total intensity of the radiation emitted by the oscillator is, according to (27.13), given by the formula

$$\bar{I} = \int d\bar{I} = \frac{e^2 \omega_0^4 r_0^2}{3c^3}. \quad (28.10)$$

The oscillator emits electromagnetic waves whose frequency is the same as its natural frequency ω_0 . The intensity of radiation is proportional to the square of the amplitude of motion and the fourth power of the frequency. The particular importance of this example is in the following.

At the beginning of the development of electron theory the well-known Thomson atomic model was proposed. It was assumed that the electron was located in the centre of a sphere formed by continuously distributed positive charge. The radiation by an atom in Thomson's model was associated with small oscillations of the electron about its equilibrium position in the centre of the atom. Thus, the atom as a radiating system was reduced to an emitting oscillator, and formula (28.10) gave the intensity of atomic radiation. In 1911 Rutherford's experiments showed Thomson's model to be incorrect, and it was abandoned. However, it turned out that the harmonic oscillator as a model of a radiating atomic system led in a number of cases to quite correct results which found experimental confirmation. The most important of these is the existence of definite frequencies ω_0 of radiation, characteristic of a given atom. Hence the oscillator remained in classical physics as a model of

a radiating atomic system, although it could not be understood within the framework of classical theory why such a model, so far from reality, could give correctly the most important properties of atomic emitters. The situation was elucidated by the appearance of the quantum theory of radiation. In quantum mechanics we shall see that the quantum theory of radiation leads, in a number of cases, to relations which are formally the same as the expressions obtained for the classical model of the emitter. This results, roughly speaking, from the following. A number of properties of atomic emitters is determined not by some concrete law of the motion of the emitting particles, but simply by the periodicity of the process. On the other hand, the circular periodic motion of an electron with a constant angular velocity corresponds to the oscillation of a plane oscillator

$$x = a \cos(\omega_0 t + \alpha), \quad y = a \sin(\omega_0 t + \alpha).$$

Hence the model of the oscillator oscillating with the frequency ω_0 yields some characteristic features of the atomic emitter. We shall take into account this fact and, later on, discuss in detail the properties of the oscillator as a classical model of an atomic radiating system. On the other hand, we shall see in a number of examples that classical electrodynamics is not directly applicable to intraatomic processes and that it leads to relations which are quantitatively, and even qualitatively, in contradiction with experimental data.

Let us now consider a system consisting of two particles with charges e_1 and e_2 and masses m_1 and m_2 . For such a system

$$\ddot{\mathbf{d}} = e_1 \ddot{\mathbf{r}}_1 + e_2 \ddot{\mathbf{r}}_2 = e_1 \mathbf{w}_1 + e_2 \mathbf{w}_2.$$

If the system of two particles is closed, the accelerations can be written in the form

$$\mathbf{w}_1 = \mathbf{F}/m_1, \quad \mathbf{w}_2 = -\mathbf{F}/m_2,$$

where \mathbf{F} is the force of interaction between the particles. Hence

$$\ddot{\mathbf{d}} = \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right) \mathbf{F}$$

and the power of radiation in the angle $d\Omega$ is

$$dI = \frac{1}{4\pi c^3} \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right)^2 [\mathbf{F} \times \mathbf{n}]^2 d\Omega. \quad (28.11)$$

The most important consequence of formula (28.11) is the statement that a closed system consisting of identical particles, or a closed system of different particles which have the same ratio e/m , cannot emit radiation in the dipole approximation. In order to find the radiation of such systems it is necessary to take into account effects of higher order (see §32).

§ 29. Radiation reaction

We have seen in preceding paragraphs that a single charge, moving with an acceleration, loses energy by radiation. In making the energy balance of a particle moving under the action of external forces it is necessary to take into account radiation losses.

We have also seen that the radiation field possesses not only energy but also momentum. Owing to this the emission of radiation is accompanied by a reaction force of the field on the particle. This action of the radiation field on the motion of the particle is called radiation reaction.

The balance of forces taking into account the radiation effect must be written in the form

$$m\mathbf{w} = \mathbf{F} + \mathbf{F}_s = \int \rho \left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{H}}{c} \right) dV + \mathbf{F}_s.$$

The first term represents the external force acting on the particle, and \mathbf{F}_s represents the radiation reaction force, called also the Lorentz deceleration force.

To calculate the radiation reaction force \mathbf{F}_s we can, in principle, proceed in the following way. We assume that the radiating charge is distributed in space. Dividing it into elements δe and $\delta e'$ we can calculate the action of the field emitted by an element $\delta e'$ on an element δe . Summing then over all elements $\delta e'$ and δe we find the total self-force sought. The calculation described appears to be rather cumbersome*. Moreover, it can be carried out only in the case of a model whose validity is, from the standpoint of contemporary quantum theory, rather weak. Therefore we shall concentrate on a different line of reasoning leading to the same expression for the force \mathbf{F}_s .

We assume that the Lorentz deceleration force \mathbf{F}_s is small in comparison with external forces. The meaning of such an assumption will be discussed in detail below, when we find \mathbf{F}_s . If this assumption is fulfilled, then in the first

* W.Heitler, *Quantum theory of radiation* (Clarendon Press, Oxford, 1954).

approximation the charge performs its motion under the action of external field forces \mathbf{F} . During this motion it radiates energy as given by formula (27.9).

In addition, we assume that the charge performs a periodic motion or, more generally, that at a certain time t_1 it comes back to the initial state of motion which it was in at the initial time t_0 . We balance the energy for a system consisting of a charge performing such a motion and an external electromagnetic field. It is obvious that, if the charge emitted nothing, then upon

its coming back into the initial state the total work $W = \int_{t_0}^{t_1} \mathbf{F} \cdot \mathbf{v} dt$ done on it

by the external field would be equal to zero. Also the change $\Delta E_{\text{ext. field}}$ in the energy of the external field would be equal to zero.

If in the next approximation we take into account that the total force acting on the charge is made up of the forces \mathbf{F} and \mathbf{F}_s , then the energy balance can be written in the form

$$\int_{t_0}^{t_1} (\mathbf{F} + \mathbf{F}_s) \cdot \mathbf{v} dt = \Delta E + \Delta E_{\text{ext. field}},$$

where ΔE is the energy emitted by the charge during the time $(t_1 - t_0)$. Taking into account that

$$\Delta E_{\text{ext. field}} = \int_{t_0}^{t_1} \mathbf{F} \cdot \mathbf{v} dt = 0,$$

it can be written that

$$\int_{t_0}^{t_1} \mathbf{F}_s \cdot \mathbf{v} dt = \Delta E = -\frac{2}{3} \frac{e^2}{c^3} \int_{t_0}^{t_1} w^2 dt. \quad (29.1)$$

Integrating the right-hand side by parts, we obtain

$$\int_{t_0}^{t_1} \mathbf{F}_s \cdot \mathbf{v} dt = -\frac{2}{3} \frac{e^2}{c^3} \mathbf{w} \cdot \mathbf{v} \Big|_{t_0}^{t_1} + \frac{2}{3} \frac{e^2}{c^3} \int_{t_0}^{t_1} \mathbf{v} \cdot \ddot{\mathbf{v}} dt.$$

Since at time t_1 the state of motion is the same as that at time t_0 , so that $\mathbf{v}_{t_0} = \mathbf{v}_{t_1}$, $\mathbf{w}_{t_0} = \mathbf{w}_{t_1}$, we have

$$\int_{t_0}^{t_1} \mathbf{F}_s \cdot \mathbf{v} dt = \frac{2}{3} \frac{e^2}{c^3} \int_{t_0}^{t_1} \mathbf{v} \cdot \ddot{\mathbf{v}} dt.$$

Equating the integrands, we find

$$\mathbf{F}_s = \frac{2}{3} \frac{e^2}{c^3} \ddot{\mathbf{v}} = \frac{2}{3} \frac{e^2}{c^3} \mathbf{w}. \quad (29.2)$$

The radiation reaction force turns out to be dependent on the derivative of the acceleration of the particle. According to the assumption made that the radiation reaction force \mathbf{F}_s is small in comparison with the external force acting on the particle, the \mathbf{w} in (29.1) should be understood to be the acceleration of the particle in the external force field. If it were not so, i.e. if, for example, the reverse inequality $\mathbf{F}_s \gg \mathbf{F}$ were fulfilled, then the equation of motion would have the form

$$m\mathbf{w} \approx \mathbf{F}_s = \frac{2}{3} \frac{e^2}{c^3} \dot{\mathbf{w}}.$$

Its solution is

$$\mathbf{w} = \mathbf{w}_0 \exp \frac{3mc^3}{2e^2} t.$$

This last formula shows that under the action of the reverse radiation reaction the acceleration increases exponentially in time — the particle is self-accelerated. Such a self-acceleration contradicts the laws of classical mechanics as well as all experimental data.

Thus, the assumption that $|\mathbf{F}_s| \gg |\mathbf{F}|$ leads to a physically senseless result. On the contrary, for $|\mathbf{F}_s| \ll |\mathbf{F}|$ the quantity \mathbf{w} can, with a sufficient degree of accuracy, be considered as equal to the acceleration acquired by the particle in the external field under the action of the Lorentz force \mathbf{F} . Assuming the latter to be a periodic function of time with a frequency ω we can, obviously, write the following expression for the amplitude of \mathbf{F}_s :

$$|\mathbf{F}_s| = \frac{2}{3} \frac{e^2}{c^3} |\dot{\mathbf{w}}| = \frac{2}{3} \frac{e^2}{c^3} \left| \frac{\dot{\mathbf{F}}}{m} \right| = \frac{2}{3} \frac{e^2}{c^3} \frac{\omega}{m} |\mathbf{F}|. \quad (29.3)$$

Hence the condition of applicability of expression (29.3) for the radiation reaction $|F_s| \ll |F|$ assumes the following form:

$$e^2 \omega / mc^3 \ll 1 \quad (29.4)$$

Violation of the inequality (29.4) means that the radiation reaction is not small, and the latter leads to a physically incorrect result. Thus, the inequality (29.4) is of fundamental importance for the applicability of radiation theory and the laws of classical field theory in general. The classical field theory only leads to reasonable results which agree with experimental data for the frequencies

$$\omega \ll mc^3 / e^2 \quad (29.5)$$

If m is understood to be the mass of the elementary charge — an electron — then the condition (29.5) turns out to be fulfilled at all optical and X-ray frequencies and even for not too high-energy γ -rays. However, as will be explained in more detail in §17 of Part II, in the case of hard γ -rays the laws of classical electrodynamics turn out to be inapplicable and quantum effects play the basic role.

It is of interest to rewrite the inequality (29.5), introducing the wavelength $\lambda = 2\pi c / \omega$ into it. Then instead of (29.5) we have

$$\lambda \gg e^2 / mc^2 = r_0 \quad (29.6)$$

For reasons which will be explained in §13 of Part II, the quantity $r_0 \approx 2.5 \times 10^{-13}$ cm is called the classical radius of the electron.

The inequality (29.6) can be interpreted as follows. For electromagnetic phenomena on the scale $\sim r_0$ the radiation reaction is not small in comparison with other forces. In this case the relations of classical field theory turn out to be inapplicable. Thus, in classical field theory there is an intrinsic limit of applicability; it is suitable for the consideration of phenomena taking place in a region of space greater than the order of the classical radius of the electron. In what follows we shall see that the actual region of applicability of classical field theory does not extend to such small dimensions. It turns out that quantum effects, which impose a limit upon the applicability of classical concepts to microparticles, begin to play a role at distances $\sim \Lambda = \hbar / mc$, where \hbar is the Planck constant equal to 1.05×10^{-27} erg·sec. The quantity Λ , equal to 2×10^{-10} cm for the electron, is called the Compton wavelength. We shall encounter this quantity in §17 of Part II, where the so-called Compton effect will be considered.

In quantum mechanics we shall discuss in detail the question of the range of applicability of the relations of classical theory, and give the proof of the statement presented.

§30. Line width of emitted radiation

The radiation reaction has a fundamental effect on the properties of the field emitted. We shall show that an emitter which in the absence of the reaction force would emit monochromatic electromagnetic waves with a frequency ω_0 in reality emits a range of frequencies in the region of ω_0 .

In other words, because of the damping effect the monochromatic radiation becomes radiation with a continuous spectrum of waves of all possible frequencies.

We shall present the proof of this statement for the simplest model of an emitting system — the linear harmonic oscillator. Let a charged particle move under the action of a quasi-elastic force $(-kx)$ along the x -axis, so that its equation of motion has the form

$$\ddot{x} = -\omega_0^2 x + \frac{F_s}{m},$$

or

$$\ddot{x} = -\omega_0^2 x + \frac{2}{3} \frac{e^2 \dot{w}}{mc^3},$$

where $\omega_0 = \sqrt{k/m}$ is the frequency of the oscillator in the absence of the force F_s .

Considering F_s to be small in comparison with the quasi-elastic force $(-kx)$, the acceleration can be assumed to be equal to the acceleration of the harmonic oscillator without radiation reaction, i.e.

$$\dot{w} = -\omega_0^2 x, \quad \dot{w} = -\omega_0^2 \dot{x},$$

and it can be written that

$$\ddot{x} + \gamma \dot{x} + \omega_0^2 x = 0, \quad (30.1)$$

where

$$\gamma = \frac{2}{3} \frac{e^2 \omega_0^2}{mc^3}. \quad (30.2)$$

The expression

$$x \approx x_0 e^{-\frac{1}{2}\gamma t} e^{i\omega_0 t} \quad (30.3)$$

serves as the solution of eq. (30.1) for $\gamma \ll \omega_0$. In this case it is to be understood that in (30.3) and subsequent formulae containing complex expressions one has to take the real part. The solution (30.3) corresponds to the initial conditions

$$x(0) = x_0, \quad \dot{x}(0) = 0.$$

Formula (30.3) shows that, because of the radiation reaction determined by the quantity γ , the oscillations are damped. The damping coefficient $\frac{1}{2}\gamma$ is analogous to the damping coefficient of the mechanical oscillator in the presence of a frictional force. This justifies the second name of the force F_s — the Lorentz friction force.

To find the radiation of a damped oscillator we write its acceleration in the form

$$w = \ddot{x} = A e^{-\frac{1}{2}\gamma t} e^{i\omega_0 t},$$

where A is a constant (in the same approximation ($\gamma \ll \omega_0$), $A \approx -x_0\omega_0^2$).

The acceleration of a damped oscillator is not a periodic function of time. Hence electromagnetic waves emitted by such an oscillator have no definite frequency. On the contrary, all frequencies $0 \leq \omega < \infty$ will be present in the radiation. This means that a damped oscillator emits a continuous spectrum of frequencies.

In what follows we shall be interested in the energy distribution in this spectrum, i.e. in the fraction of the total energy emitted by the oscillator per frequency interval $\omega, \omega + d\omega$. This function $I(\omega)$, called the Lorentz spectral distribution function, is related to the total energy I_0 emitted by the oscillator:

$$I_0 = \int_0^\infty I(\omega) d\omega. \quad (30.4)$$

The total energy emitted by the oscillator is

$$I_0 = \int_0^\infty I dt = \frac{2}{3} \frac{e^2}{c^3} \int_0^\infty w^2 dt = \frac{2}{3} \frac{e^2}{c^3} \int_{-\infty}^\infty w^2 dt \quad (30.5)$$

Here we extend the integration range to the region of negative times, since at $t < 0$ the oscillator was at rest and the integrand is identically equal to zero.

We expand the acceleration in a Fourier integral:

$$w(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} W(\omega) e^{i\omega t} d\omega,$$

where the Fourier component $W(\omega)$ is equal to

$$\begin{aligned} W(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} w(t) e^{-i\omega t} dt = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} w(t) e^{-i\omega t} dt = \\ &= \frac{A}{\sqrt{2\pi}} \frac{1}{[\frac{1}{2}\gamma - i(\omega_0 - \omega)]}. \end{aligned}$$

According to formula (II.9) we have

$$\int_{-\infty}^{\infty} w^2 dt = \int_{-\infty}^{\infty} |W(\omega)|^2 d\omega = \frac{A^2}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega}{(\omega_0 - \omega)^2 + \frac{1}{4}\gamma^2} = \frac{A^2}{\gamma} \quad (30.6)$$

Substituting formula (30.6) into (30.5), we find

$$I_0 = \frac{1}{3\pi} \frac{e^2}{c^3} A^2 \int_{-\infty}^{\infty} \frac{d\omega}{(\omega_0 - \omega)^2 + \frac{1}{4}\gamma^2} = \frac{2}{3} \frac{e^2}{c^3} \frac{A^2}{\gamma}, \quad (30.7)$$

whence

$$A^2 = \frac{3c^3\gamma}{2e^2} I_0. \quad (30.8)$$

On the other hand, comparing (30.7) and (30.4) and taking into account that the spectral distribution is determined only for essentially positive values of the frequencies, we find

$$I(\omega) = \frac{I_0}{2\pi} \frac{\gamma}{[(\omega_0 - \omega)^2 + \frac{1}{4}\gamma^2]}. \quad (30.9)$$

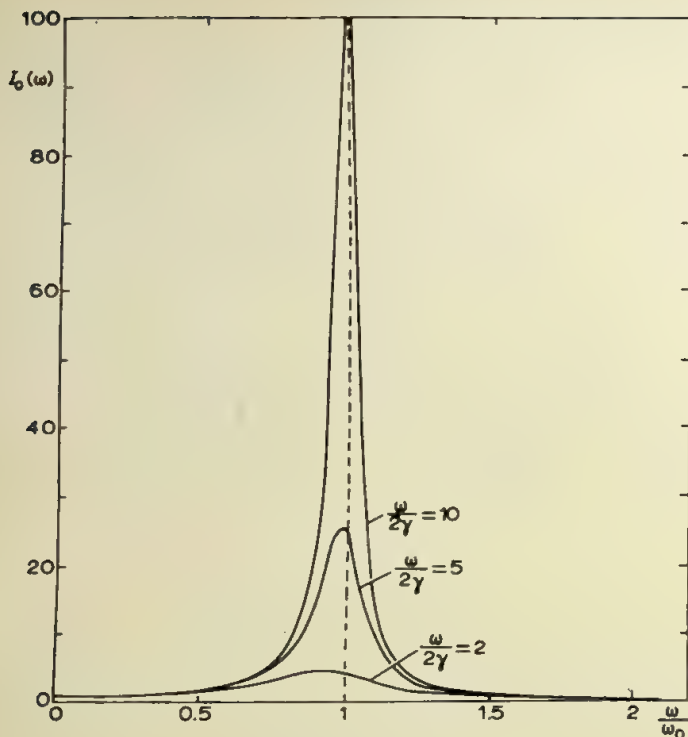


Fig. 1.11

The spectral distribution functions for different values of $\omega_0/2\gamma$ are shown in fig. 1.11. They have a sharp maximum for $\omega \approx \omega_0$, i.e. for the frequency which would be emitted by the oscillator in the absence of damping,

$$I(\omega_0) = 2I_0/\pi\gamma.$$

For $\omega = \omega_0 \pm \frac{1}{2}\gamma$ the emitted intensity is equal to

$$I(\omega_0 \pm \frac{1}{2}\gamma) = \frac{1}{2}I(\omega_0),$$

i.e. is lower by a factor of two than the intensity in the maximum. For this reason the quantity $\frac{1}{2}\gamma$ is called the half-width of the emitted line. According to (30.2), there corresponds to the line half-width $\frac{1}{2}\gamma$, the wavelength interval

$$\begin{aligned}
 |\Delta\lambda| &= 2\pi \left| \Delta \frac{c}{\omega_0} \right| = \frac{2\pi c \Delta\omega}{\omega_0^2} = \frac{2\pi c}{\omega_0^2} \frac{\gamma}{2} \\
 &= \frac{\pi c}{\omega_0^2} \frac{2}{3} \frac{e^2}{c^3 m} \omega_0^2 = \frac{2\pi}{3} \frac{e^2}{mc^2} = \frac{2\pi}{3} r_0,
 \end{aligned}$$

which does not depend on the wavelength and whose order of magnitude is equal to that of the classical radius of the electron.

§31. Quadrupole and magnetic dipole radiation

We have already seen above that in certain cases a system of charges cannot emit radiation in the dipole approximation. This does not mean, of course, that such a system cannot emit radiation at all.

If there is no radiation in the dipole approximation, one has to look for higher terms of the expansion in powers of the proper retardation in the system, which will determine a radiation of higher order — quadrupole radiation, octupole radiation and so on. We shall restrict ourselves to finding the radiation of the next order beyond the dipole approximation.

We write the vector potential of the emitting system, at a large distance from the system, in a form analogous to (26.5):

$$\begin{aligned}
 \mathbf{A}(\mathbf{r}, t) &= \frac{1}{c} \int \frac{1}{r} \left[\mathbf{j}(\mathbf{r}', \tau_0) + \frac{\mathbf{r}' \cdot \mathbf{r}}{cr} \frac{\partial \mathbf{j}}{\partial \tau_0} \right] dV' = \\
 &= \frac{1}{cr} \int \mathbf{j}(\mathbf{r}', \tau_0) dV' + \frac{1}{c^2 r} \int \frac{\partial \mathbf{j}}{\partial \tau_0} (\mathbf{n} \cdot \mathbf{r}') dV' = \mathbf{A}_1 + \mathbf{A}_2.
 \end{aligned} \tag{31.1}$$

According to (26.9), the first term of (31.1) describes the dipole radiation. Hence we shall be interested only in the second term:

$$\mathbf{A}_2 = \frac{1}{c^2 r} \int \frac{\partial \mathbf{j}(\mathbf{r}', \tau_0)}{\partial \tau_0} (\mathbf{n} \cdot \mathbf{r}') dV' = \frac{1}{c^2 r} \frac{\partial}{\partial \tau_0} \int \mathbf{j}(\mathbf{n} \cdot \mathbf{r}') dV' \tag{31.2}$$

Here we have made use of the fact that the constant vector \mathbf{n} and the integration constant \mathbf{r}' do not depend on time, and we have changed the order of the differentiation and integration. The value of the integral is taken at time τ_0 .

We transform the integrand, symmetrizing it by means of formula (1.6).

We then obtain

$$\begin{aligned} \mathbf{A}_2 = & \frac{1}{2c^2 r} \frac{\partial}{\partial \tau_0} \left\{ \int (\mathbf{r}' \times \mathbf{j}) \times \mathbf{n} dV' \right\} \\ & + \frac{1}{2c^2 r} \frac{\partial}{\partial \tau_0} \int \{ \mathbf{j}(\mathbf{r}' \cdot \mathbf{n}) + \mathbf{r}'(\mathbf{j} \cdot \mathbf{n}) \} dV'. \end{aligned} \quad (31.3)$$

From the definition of the magnetic moment (22.1) it follows that the first integral can be written in the form

$$\frac{1}{2c} \int (\mathbf{r}' \times \mathbf{j}) \times \mathbf{n} dV' = \mathbf{M} \times \mathbf{n}.$$

In order to find the meaning of the second integral, it is convenient to go over from integration to summation *, projecting the vector expression onto a particular axis α in the Cartesian coordinate frame:

$$\begin{aligned} \int \{ \mathbf{j}(\mathbf{r}' \cdot \mathbf{n}) + \mathbf{r}'(\mathbf{j} \cdot \mathbf{n}) \} \alpha dV' &= \sum_i \{ e_i \mathbf{v}_i (\mathbf{r}'_i \cdot \mathbf{n}) + e_i \mathbf{r}'_i (\mathbf{v}_i \cdot \mathbf{n}) \} \alpha = \\ &= \frac{\partial}{\partial \tau_0} \left\{ \sum e \mathbf{r}'_i (\mathbf{r}'_i \cdot \mathbf{n}) \right\} \alpha = \frac{1}{3} \frac{\partial}{\partial \tau_0} \{ D_{\alpha\beta} n_\beta \}, \end{aligned}$$

where $D_{\alpha\beta}$ is the quadrupole moment of the system, determined by formula (16.3). The summation is carried out with respect to the index β ($\beta = x, y, z$). Going over to the vector expression, we can write

* We draw attention to the transition from integration to summation. For a simpler expression it can be written in more detail:

$$\begin{aligned} \int \mathbf{r}' \cdot \frac{\partial \mathbf{j}}{\partial t} dV' &= \frac{\partial}{\partial t} \int \mathbf{r}' \cdot \mathbf{j} dV' = \frac{\partial}{\partial t} \int \mathbf{r}' \cdot \sum e \mathbf{v}_i \delta(\mathbf{r}' - \mathbf{r}_i) dV' = \\ &= \frac{\partial}{\partial t} \sum e_i \mathbf{v}_i \cdot \mathbf{r}_i = \sum e_i \left(\frac{d\mathbf{v}_i}{dt} \cdot \mathbf{r}_i + \mathbf{v}_i \cdot \frac{d\mathbf{r}_i}{dt} \right). \end{aligned}$$

The integration variable \mathbf{r}' is a constant in the differentiation. Only the coordinate of the i th charge \mathbf{r}_i depends on time. Hence the order of the differentiation with respect to t and integration with respect to \mathbf{r}' can be changed. However, after going over to the summation it is necessary to differentiate \mathbf{r}_i .

$$\int \{ \mathbf{j}(\mathbf{r}' \cdot \mathbf{n}) + \mathbf{r}'(\mathbf{j} \cdot \mathbf{n}) \} dV' = \frac{\partial}{\partial \tau_0} \mathbf{D}, \quad (31.4)$$

where the vector \mathbf{D} is defined by

$$D_\alpha = D_{\alpha\beta} n_\beta. \quad (31.5)$$

Then, finally, we find for the vector potential

$$\mathbf{A} = \frac{\dot{\mathbf{d}}}{cr} + \frac{\dot{\mathbf{M}} \times \mathbf{n}}{cr} + \frac{\ddot{\mathbf{D}}}{6c^2 r} = \mathbf{A}_{\text{dip}} + \mathbf{A}_{\text{magn. dip}} + \mathbf{A}_{\text{quadr}}, \quad (31.6)$$

where the dot denotes the differentiation with respect to the argument τ_0 on which the vectors \mathbf{d} , \mathbf{M} and \mathbf{D} depend.

In the expression (31.6) the first term describes dipole radiation. The second term is determined by the derivative with respect to time of the magnetic moment and, naturally, is called magnetic dipole radiation. The last term contains the second derivative of the quadrupole moment. It determines the quadrupole radiation.

We estimate the order of magnitude of the terms of (31.6):

$$\dot{\mathbf{d}} \sim eL'\omega, \quad \dot{\mathbf{M}} \sim \frac{e}{c} vL'\omega \approx \frac{eL'^2\omega^2}{c}, \quad \ddot{\mathbf{D}} \sim eL'^2\omega^2.$$

Hence

$$\frac{|\dot{\mathbf{M}} \times \mathbf{n}|}{|\dot{\mathbf{d}}|} \sim \frac{evL'}{ecL'} \sim \frac{v}{c},$$

$$\frac{1}{c} \frac{|\ddot{\mathbf{D}}_{\alpha\beta} n_\beta|}{|\dot{\mathbf{d}}|} \sim \frac{e\omega^2 L'^2}{ec\omega L'} \sim \frac{L'\omega}{c} \sim \frac{L'}{\lambda},$$

where L' is the characteristic dimension of the system, and λ is the wavelength of the radiation.

Since according to our assumptions $v \ll c$ and $L' \ll \lambda$, the terms corresponding to magnetic dipole radiation and quadrupole radiation are very small in comparison with the first term describing dipole radiation. This means that magnetic dipole radiation and quadrupole radiation are significant only for systems for which the electric dipole radiation is absent. In general it cannot be said which of these two terms then gives the main contribution to the radiation.

We shall calculate separately the intensity of the magnetic dipole radiation and the quadrupole radiation.

According to formulae (27.8) and (27.1), the intensity of the magnetic dipole radiation is

$$\begin{aligned} dI_m &= \frac{c}{4\pi} H^2 r^2 d\Omega = \frac{1}{4\pi c} [\dot{\mathbf{A}} \times \mathbf{n}]^2 r^2 d\Omega = \\ &= \frac{1}{4\pi c^3} [(\ddot{\mathbf{M}} \times \mathbf{n}) \times \mathbf{n}]^2 d\Omega = \frac{1}{4\pi c^3} \ddot{M}^2 \sin^2 \theta d\theta d\psi. \quad (31.7) \end{aligned}$$

The intensity of the magnetic dipole radiation has exactly the same form as the intensity of the electric dipole radiation (27.8), but in (31.7) we have $\ddot{\mathbf{M}}$ instead of $\ddot{\mathbf{d}}$.

The total intensity of the magnetic dipole radiation is obtained by integrating (31.7) over all angles:

$$I_m = \frac{2}{3} \frac{(\ddot{\mathbf{M}})^2}{c^3}. \quad (31.8)$$

The intensity of the magnetic dipole radiation is smaller than that of the electric dipole radiation by the ratio $(v/c)^2$.

It should be stressed that magnetic dipole radiation is absent for systems whose magnetic moment is proportional to the angular momentum. According to what was said in §22, this holds for systems with a constant value of the ratio e/m , and also for systems of two arbitrary particles. Because of this, magnetic dipole radiation is absent in collisions between two particles.

We now go on to the calculation of the intensity of the quadrupole radiation. Substituting the value of $\mathbf{A}_{\text{quadr}}$ into the general formula (27.6) and reproducing the calculations of §27, we find

$$\begin{aligned} dI_{\text{quadr}} &= \frac{c}{4\pi} H^2 r^2 d\Omega = \frac{1}{4\pi c} [\dot{\mathbf{A}}_{\text{quadr}} \times \mathbf{n}]^2 r^2 d\Omega = \\ &= \frac{1}{4\pi c} \frac{1}{36c^4} [\ddot{\mathbf{D}} \times \mathbf{n}]^2 d\Omega. \quad (31.9) \end{aligned}$$

We transform the square of the vector product according to formula (I.5) taking into account the definition (31.5):

$$[\ddot{\mathbf{D}} \times \mathbf{n}]^2 = (\ddot{\mathbf{D}})^2 - (\mathbf{n} \cdot \ddot{\mathbf{D}})^2 = \ddot{D}_{\alpha\beta} \ddot{D}_{\alpha\gamma} n_\beta n_\gamma - \ddot{D}_{\alpha\beta} \ddot{D}_{\lambda\mu} n_\alpha n_\beta n_\lambda n_\mu. \quad (31.10)$$

Summation over repeated indices is implied.

The angular dependence of dI_{quadr} turns out to be very complex. However, the total radiation per unit time can be calculated relatively simply. Integrating (31.9), taking into account (31.10), we obtain

$$I_{\text{quadr}} = \frac{1}{4\pi c} \frac{1}{36c^4} \left\{ \ddot{D}_{\alpha\beta} \ddot{D}_{\alpha\gamma} \int n_\beta n_\gamma d\Omega - \ddot{D}_{\alpha\beta} \ddot{D}_{\lambda\mu} \int n_\alpha n_\beta n_\lambda n_\mu d\Omega \right\} \quad (31.11)$$

The calculation gives:

$$\int n_\beta n_\gamma d\Omega = \frac{4}{3} \pi \delta_{\beta\gamma}, \quad (31.12)$$

$$\int n_\alpha n_\beta n_\lambda n_\mu d\Omega = \frac{4}{15} \pi (\delta_{\alpha\beta} \delta_{\lambda\mu} + \delta_{\alpha\lambda} \delta_{\beta\mu} + \delta_{\alpha\mu} \delta_{\beta\lambda}). \quad (31.13)$$

Hence for I we find

$$\begin{aligned} I_{\text{quadr}} &= \frac{1}{36c^5} \left\{ \frac{1}{3} (\ddot{D}_{\alpha\beta})^2 - \frac{1}{15} (\ddot{D}_{\alpha\alpha} \ddot{D}_{\beta\beta} + 2\ddot{D}_{\alpha\beta}^2) \right\} = \\ &= \frac{1}{180c^5} (\ddot{D}_{\alpha\beta})^2, \end{aligned} \quad (31.14)$$

since, according to (16.6), the sum of the diagonal elements $\ddot{D}_{\alpha\alpha}$ is always equal to zero. When the quadrupole moment varies periodically, the quadrupole radiation turns out to be proportional to ω^6 . The factor $1/c^5$ in (31.14) makes its intensity very small. Nevertheless, there are important examples of systems for which quadrupole radiation plays the basic role. These are, first of all, some atomic nuclei emitting no electric dipole radiation (this is due to the law of conservation of angular momentum; for a justification of this assertion see Part V, §106).

A closed system consisting of particles with the same values of e/m could serve as another example (see eq. (28.11)).

§32*. General case of electromagnetic radiation. The spectral decomposition of fields. The radiation zone and induction zone. Effect of the proper retardation

We shall now consider a somewhat more general case of radiation emitted by a system of moving charges.

Without restricting the general character of the treatment it can be assumed that the current density in the system and the charge density may be expanded in Fourier integrals, so that they can be written

$$\mathbf{j}(\mathbf{r}', t) = \int \mathbf{j}(\mathbf{r}', \omega) e^{i\omega t} d\omega, \quad (32.1)$$

$$\rho(\mathbf{r}', t) = \int \rho(\mathbf{r}', \omega) e^{i\omega t} d\omega. \quad (32.2)$$

In these expansions we make use of the normalization (II.1) and assume formally that the frequency may be negative as well as positive. In reality the frequency ω is a quantity which is essentially positive, so that this representation should be supplemented by the condition $\mathbf{j}(\mathbf{r}', \omega) = \mathbf{j}(\mathbf{r}', -\omega)$. The quantity $\mathbf{j}(\mathbf{r}', t)$ will have a real value. Then, expanding the field potential in a Fourier integral

$$A(\mathbf{r}, t) = \int A(\mathbf{r}, \omega) e^{i\omega t} d\omega, \quad (32.3)$$

and substituting the expansions (32.1)–(32.3) into eq. (23.23) for the vector potential, we find

$$\int A(\mathbf{r}, \omega) e^{i\omega t} d\omega = \frac{1}{c} \int \frac{\mathbf{j}(\mathbf{r}', \omega)}{|\mathbf{r} - \mathbf{r}'|} \exp \left\{ i\omega \left(t - \frac{|\mathbf{r} - \mathbf{r}'|}{c} \right) \right\} d\omega dV$$

Whence, for the Fourier component of the vector potential, we obtain

$$A(\mathbf{r}, \omega) = \frac{1}{c} \int \frac{\mathbf{j}(\mathbf{r}', \omega)}{|\mathbf{r} - \mathbf{r}'|} \exp \left\{ -i \frac{\omega}{c} |\mathbf{r} - \mathbf{r}'| \right\} dV' \quad (32.4)$$

Analogously,

$$\varphi(\mathbf{r}, \omega) = \int \frac{\rho(\mathbf{r}', \omega)}{|\mathbf{r} - \mathbf{r}'|} \exp \left\{ -i \frac{\omega}{c} |\mathbf{r} - \mathbf{r}'| \right\} dV' \quad (32.5)$$

If the current density and charge density are given as functions of coordinates, then from formulae (32.4) and (32.5) one can find the Fourier components

of the field potentials accurately, without disregarding the proper retardation in the system.

Hence the expressions for the potentials themselves can be found directly from the formula (II.2) of the Fourier transform.

Thus, it is in principle possible to find the field of an emitting system without disregarding the proper retardation, and at any distance from the emitting system. Below we shall present a simple example of such a calculation.

If we are interested in the field at large distances from the emitting system, then, assuming

$$|\mathbf{r} - \mathbf{r}'| \approx r - \frac{\mathbf{r} \cdot \mathbf{r}'}{r} = \frac{\mathbf{r}}{r} \cdot (\mathbf{r} - \mathbf{r}'),$$

where $|\mathbf{r}| \gg |\mathbf{r}'|$, and introducing a quantity called the wave vector,

$$\mathbf{k} = \frac{\omega}{c} \frac{\mathbf{r}}{r} = \frac{\omega}{c} \mathbf{n} = k \mathbf{n}, \quad (32.6)$$

where k is the wave number

$$k = \frac{\omega}{c} = \frac{2\pi}{\lambda} \quad (32.7)$$

and \mathbf{n} is a unit vector, we have

$$\mathbf{A}(\mathbf{r}, \omega) = \frac{e^{-ikr}}{cr} \int \mathbf{j}(\mathbf{r}', \omega) e^{i\mathbf{k} \cdot \mathbf{r}'} dV', \quad (32.8)$$

$$\varphi(\mathbf{r}, \omega) = \frac{e^{-ikr}}{r} \int \rho(\mathbf{r}', \omega) e^{i\mathbf{k} \cdot \mathbf{r}'} dV', \quad (32.9)$$

when in the denominator the approximation $|\mathbf{r} - \mathbf{r}'| \approx r$ is made. The formulae of spectral decomposition are often applied to the determination of the field of a single charge. In this case one can obtain a useful representation of $\mathbf{A}(\mathbf{r}, \omega)$ by substituting for the Fourier component of the current density the current density itself according to the formula

$$\mathbf{j}(\mathbf{r}', \omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \mathbf{j}(\mathbf{r}', t) e^{-i\omega t} dt,$$

and assuming

$$\mathbf{j}(\mathbf{r}', t) = e\mathbf{v}\delta(\mathbf{r}' - \mathbf{R}_0(t)) ,$$

where $\mathbf{R}_0(t)$ is the instantaneous position of the charge at time t , and $\mathbf{v} = d\mathbf{R}_0/dt$. This gives

$$\mathbf{A}(\mathbf{r}, \omega) = \frac{e^{-ikr}}{2\pi cr} \int_{-\infty}^{+\infty} e\mathbf{v}(t) e^{-i[\omega t - \mathbf{k} \cdot \mathbf{R}_0(t)]} dt . \quad (32.10)$$

The above formula allows one to find the Fourier components of the vector potential from the trajectory of the particle — the dependence of \mathbf{R}_0 on time. The quantity r represents the absolute value of \mathbf{r} — the distance from the origin to the point of observation at time t . If the Fourier components of the vector potential at a large distance from the system of charges are known, then the Fourier components of the field can be found from formulae (27.1) and (27.3). Namely, we have

$$\begin{aligned} \mathbf{H}(\mathbf{r}, t) &= \int \mathbf{H}(\mathbf{r}, \omega) e^{i\omega t} d\omega = \nabla \times \mathbf{A}(\mathbf{r}, t) = \\ &= \frac{1}{c} \dot{\mathbf{A}} \times \mathbf{n} = \frac{1}{c} \int i\omega [\mathbf{A}(\mathbf{r}, \omega) \times \mathbf{n}] e^{i\omega t} d\omega . \end{aligned}$$

Whence

$$\mathbf{H}(\mathbf{r}, \omega) = i[\mathbf{A}(\mathbf{r}, \omega) \times \mathbf{k}] . \quad (32.11)$$

Analogously,

$$\mathbf{E} = \frac{i}{k} [(\mathbf{A}(\mathbf{r}, \omega) \times \mathbf{k}) \times \mathbf{k}] . \quad (32.12)$$

In general a radiating system emits all possible frequencies or, as is usually said, a frequency spectrum $-\infty < \omega < \infty$.

It is often of interest to find the weight of different frequencies in a spectrum, i.e. the relative fraction of energy for a given frequency (or, more precisely, for a frequency lying in an interval $\omega, \omega + d\omega$). We write the total energy radiated into a solid angle $d\Omega$ in a time dt in the form

$$\Delta E = \int I dt = \frac{c}{4\pi} \int H^2 r^2 d\Omega dt = d\Omega \int J_\omega d\omega dt , \quad (32.13)$$

where $J_\omega d\omega d\Omega$ is the power radiated into a solid angle $d\Omega$ in a frequency interval $d\omega$.

Expanding \mathbf{H} in a Fourier integral and making use of the Parseval equality (II.9), we have

$$\int (H(\mathbf{r}, t))^2 dt = 4\pi \int (H(\mathbf{r}, \omega))^2 d\omega.$$

Then for the power (energy per unit time) of the radiation in a solid angle $d\Omega$ in a frequency interval $d\omega$ we obtain the expression

$$\begin{aligned} J_\omega d\Omega d\omega &= c |\mathbf{H}(\mathbf{r}, \omega)|^2 r^2 d\Omega d\omega = \\ &= \frac{1}{c} \left| [\mathbf{k} \times \int \mathbf{j}(\mathbf{r}', \omega) e^{i\mathbf{k} \cdot \mathbf{r}'} dV'] \right|^2 d\Omega d\omega. \end{aligned} \quad (32.14)$$

The actual spectral distribution of the emitted energy is, naturally, determined by the law of motion of the charges $\mathbf{j}(\mathbf{r}', t)$.

The above formulae are simplified if the current density and charge density vary at a single frequency ω according to the simple harmonic law:

$$\begin{aligned} \mathbf{j}(\mathbf{r}', t) &= \mathbf{j}_0(\mathbf{r}') e^{i\omega t}, \\ \rho(\mathbf{r}', t) &= \rho_0(\mathbf{r}') e^{i\omega t} \end{aligned} \quad (32.15)$$

In this case the system, naturally, emits only one frequency ω , and the vector potential is given using the general formula (24.16) with $d\mathbf{r}' = dV'$

$$\mathbf{A}(\mathbf{r}, t) = \frac{e^{i\omega t}}{c} \int \frac{\mathbf{j}_0(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \exp \left\{ -i \frac{\omega}{c} |\mathbf{r} - \mathbf{r}'| \right\} dV' = \mathbf{A}_0(\mathbf{r}) e^{i\omega t}, \quad (32.16)$$

where the amplitude of the potential $\mathbf{A}_0(\mathbf{r})$ is, obviously, equal to

$$\mathbf{A}_0(\mathbf{r}) = \frac{1}{c} \int \frac{\mathbf{j}_0(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \exp \{ -ik |\mathbf{r} - \mathbf{r}'| \} dV'. \quad (32.17)$$

At a large distance from the emitting system,

$$\mathbf{A}(\mathbf{r}, t) \approx \frac{e^{-i(kr - \omega t)}}{cr} \int \mathbf{j}_0(\mathbf{r}') e^{i\mathbf{k} \cdot \mathbf{r}'} dV' \quad (32.18)$$

The quantity $\mathbf{k} \cdot \mathbf{r}'$ obviously determines the retardation inside the system (the proper retardation).

Now consider the case when the proper retardation can be disregarded. For this it is necessary that the inequality

$$\mathbf{k} \cdot \mathbf{r}' \ll 1$$

or

$$L' \ll \lambda \quad (32.19)$$

be satisfied. We see that the condition for disregarding the proper retardation is the smallness of the geometrical dimensions of the emitting system in comparison with the wavelength of the radiation.

Assuming that the condition (32.19) is fulfilled, we have, by virtue of (26.11),

$$\mathbf{A}(\mathbf{r}, t) = \frac{e^{-i(kr - \omega t)}}{cr} \int \mathbf{j}_0(\mathbf{r}') dV' = \frac{\dot{\mathbf{d}}(\tau_0)}{cr}, \quad (32.20)$$

which is in agreement with (26.13).

Let us also find the scalar potential, assuming, however, that the distance to the point of observation is not large in comparison with the wavelength. In order not to repeat the calculations of §26, we write the expression for the scalar potential making use of the Lorentz condition:

$$\begin{aligned} \varphi &= -c \int \nabla \cdot \mathbf{A} dt = -c \nabla \cdot \frac{\dot{\mathbf{d}}(\tau_0)}{cr} = \\ &= \frac{\mathbf{n} \cdot \dot{\mathbf{d}}(\tau_0)}{r} \left(ik + \frac{1}{r} \right). \end{aligned} \quad (32.21)$$

We see that two limiting cases can be realized:

$$\mathbf{k} \cdot \mathbf{r} \sim r/\lambda \gg 1,$$

$$\mathbf{k} \cdot \mathbf{r} \sim r/\lambda \ll 1.$$

In the first case, when the distance from the emitter is large in comparison with its dimensions, we obtain

$$\varphi(\mathbf{r}, t) = ik \frac{\mathbf{n} \cdot \mathbf{d}(\tau_0)}{r}. \quad (32.22)$$

which is in agreement with (26.12). It should be recalled that in formula (32.22) the real part of the corresponding expression is implied, and the factor i plays no role. Thus, in the range

$$r \gg \lambda \gg L'$$

all the formulae of §26 are valid. This is the region of large distances or the radiation zone.

In the range of distances

$$r \ll \lambda \ll L'$$

we can write

$$\varphi(\mathbf{r}, t) \approx \frac{\mathbf{n} \cdot \mathbf{d}(\tau_0)}{r^2} = \frac{\mathbf{n} \cdot \mathbf{d}_0 e^{i\omega\tau_0}}{r^2}. \quad (32.23)$$

In this region, called the neighbouring or induction zone, the scalar potential φ is the same as the potential of a dipole electrostatic field, whose value varies according to a harmonic law. The vector potential \mathbf{A} in this zone is small in comparison with the scalar potential by the ratio r/λ . This means that in the induction zone the field has, basically, the character of the electric field. The strength of the electric field varies in the entire induction zone in phase with the variation of the dipole moment of the system of charges, and the field does not have a wave character.

The magnetic field is smaller than the electric field by the ratio $|\mathbf{H}|/|\mathbf{E}| \approx r/\lambda$ in contrast to the radiation zone, where the relation (27.3) is fulfilled. We see that in emitting long waves ($\lambda \gg L'$) the field of the emitting system has an induction character in the region $r \ll \lambda$, and a wave character for $r \gg \lambda$. The angular distribution of the electric field in the two regions is different. In the induction zone the energy flux averaged over the time, characterized by the Poynting vector, is equal to zero. Thus, the induction zone gives no contribution to the radiation.

We also write the expressions for the fields in monochromatic waves at a large distance from the emitting system:

$$\mathbf{H} = \frac{k^2}{r} \mathbf{n} \times \mathbf{d}_0 e^{i(\omega t - kr)}, \quad (32.24)$$

$$\mathbf{E} = \frac{k^2}{r} (\mathbf{n} \times \mathbf{d}_0) \times \mathbf{n} e^{i(\omega t - kr)}, \quad (32.25)$$

and in the vicinity of the emitting system:

$$\mathbf{E} = \frac{3\mathbf{n}(\mathbf{d}_0 \cdot \mathbf{n}) - \mathbf{d}_0}{r^3} e^{i\omega t}, \quad (32.26)$$

$$\mathbf{H} \cong 0$$

In the radiation zone the Poynting vector is determined by formula (27.6). Hence the emitted power averaged over the time is equal to

$$I = \frac{\omega^4 d_0^2}{3c^3} = \frac{cd_0^2}{3} \left(\frac{2\pi}{\lambda} \right)^4 \quad (32.27)$$

It turns out to be inversely proportional to the fourth power of the wavelength and directly proportional to the square of the amplitude of the dipole moment d_0 . Finally, we shall discuss an example of the calculation of the field of the emitter taking into account the retardation in the system. We consider a system in which the current density is expressed by the formula

$$\mathbf{j}(\mathbf{r}') = j_0 \mathbf{k}_1 \sin\left(\frac{1}{2}kL - k|z|\right) \delta(x) \delta(y) e^{i\omega t} \quad (32.28)$$

Here \mathbf{k}_1 is the unit vector along the z -axis.

Formula (32.28) has a simple meaning. In an infinitely thin line of length L , directed along the z -axis, a current having the character of a standing wave is produced. At the ends of the line at $z = \pm \frac{1}{2}L$ the current density reduces to zero. Such a system is called a linear radiator and is often used for emitting radiowaves.

Substituting (32.28) into (32.18), we obtain for the vector \mathbf{A}

$$\mathbf{A}(\mathbf{r}, t) = j_0 \frac{\mathbf{k}_1}{cr} \int_{-\frac{1}{2}L}^{+\frac{1}{2}L} \sin\left(\frac{1}{2}kL - k|z|\right) e^{ikz \cos \theta} e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})} dz,$$

where $\theta = (\mathbf{n} \cdot \mathbf{k}_1)$. On integrating we obtain

$$\mathbf{A} = \left\{ \mathbf{k}_1 j_0 \frac{kL^2}{4} \frac{e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}}{cr} \right\} F(\theta, k), \quad (32.29)$$

where the factor $F(\theta, k)$ is equal to

$$F(\theta, k) = \frac{8}{k^2 L^2} \frac{\cos(\frac{1}{2} kL \cos \theta) - \cos \frac{1}{2} kL}{\sin^2 \theta}. \quad (32.30)$$

The factor in the brackets corresponds to the emission of waves whose wavelength is $\lambda \gg \frac{1}{2} L$ (very long waves) and for which the proper retardation can be disregarded. It is obtained directly from (32.20) upon substituting formula (32.28), in which it is assumed that $\frac{1}{2} kL \rightarrow 0$.

The factor $F(\theta, k)$ characterizes the proper retardation in the system. It tends to one as $\frac{1}{2} kL \rightarrow 0$. For a finite value of kL it corresponds to a substantial change in the angular distribution of the radiation.

For $kL = \pi$ (i.e. $L = \frac{1}{2}\lambda$) the angular distribution of the radiation differs relatively little from that of a simple dipole. For $kL = m\pi$, where m is the integer number of half-wavelengths within the length of the linear emitter, there arise distributions within values of θ at which the radiation has a maximum value, the most intense maxima being nearest the polar axis.

With increasing m these intense maxima progressively approach the axis of the emitter. In the limit, as $m \rightarrow \infty$, the radiation turns out to be directed along the axis.

Electromagnetic Field in a Vacuum and Electromagnetic Wave Scattering

§33. The propagation of electromagnetic waves at a large distance from the emitter

In the preceding paragraphs the phenomenon of emission of electromagnetic waves has been studied. We can now analyse the mechanism of propagation of electromagnetic waves in a space free of charges (in vacuum).

We have seen in the preceding chapter that radiating systems emit electromagnetic waves for which the equal-phase surfaces are spheres of radius r .

We now pass over in formulae (32.24) and (32.25) to the limit, assuming the distance from the system of charges to be so large that one may disregard the difference in the direction of the vector \mathbf{r}_0 and \mathbf{n} , where \mathbf{r}_0 is the unit vector directed from the origin to the point of observation, and \mathbf{n} is, as before, the unit vector directed from the emitter to the observation point.

We can then write

$$r = \mathbf{r} \cdot \mathbf{r}_0 \approx \mathbf{r} \cdot \mathbf{n}$$

and make this substitution in the phase factor, after which formulae (32.24) and (32.25) will take the form

$$\mathbf{H} = \frac{\omega^2}{c^2} \frac{1}{r} \mathbf{n} \times \mathbf{d}_0 \exp i\omega \left(t - \frac{\mathbf{r} \cdot \mathbf{n}}{c} \right), \quad (33.1)$$

$$\mathbf{E} = \frac{\omega^2}{c^2} \frac{1}{r} (\mathbf{n} \times \mathbf{d}_0) \times \mathbf{n} \exp i\omega \left(t - \frac{\mathbf{r} \cdot \mathbf{n}}{c} \right). \quad (33.2)$$

In formulae (33.1) and (33.2) the distance from the point of observation to the origin enters in the amplitude in the form of the factor $1/r$. However, it is clear that at an infinitely large distance from the origin the variation of the function $1/r$ can be assumed to be very slow and we can set $1/r \approx \text{const}$. In this case the variation of the fields with distance is determined solely by the phase factor. Then (33.1) and (33.2) can be written in the form

$$\mathbf{E} = \mathbf{E}_0 \exp i\omega \left(t - \frac{\mathbf{r} \cdot \mathbf{n}}{c} \right), \quad (33.3)$$

$$\mathbf{H} = \mathbf{H}_0 \exp i\omega \left(t - \frac{\mathbf{r} \cdot \mathbf{n}}{c} \right), \quad (33.4)$$

where \mathbf{E}_0 and \mathbf{H}_0 are quantities which are constant in space and time.

We introduce the vector $\mathbf{k} = \omega \mathbf{n}/c$, called the wave vector. In absolute value $k = \omega/c = 2\pi/\lambda$, where λ is the wavelength, and is oriented in the direction of propagation of the electromagnetic wave. Then

$$\mathbf{E} = \mathbf{E}_0 e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}, \quad (33.5)$$

$$\mathbf{H} = \mathbf{H}_0 e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}. \quad (33.6)$$

In formulae (33.5) and (33.6) it is implied that in the final expressions the real part is to be taken. According to (27.3), the amplitudes \mathbf{E}_0 and \mathbf{H}_0 are perpendicular to each other and to the vector \mathbf{k} :

$$\mathbf{E}_0 = \mathbf{H}_0 \times \mathbf{n}. \quad (33.7)$$

The vectors \mathbf{E} and \mathbf{H} represent plane monochromatic waves: the planes $\mathbf{k} \cdot \mathbf{r} = \text{const}$ are equal-phase surfaces (surfaces of equal values of the strengths \mathbf{E} and \mathbf{H}) (fig. I.12).

Thus, we arrive at a natural result. At sufficiently large distances from the emitter spherical waves approximate to plane waves. This can be pictured in an obvious way as follows. If the emitting system is sufficiently distant, the radius of curvature of the spherical equal-phase surface of the electromagnetic wave is very large in comparison with the dimensions of that region of space in which the field is considered. Hence in the limits of this region the sphere can be assumed with a sufficient degree of accuracy to be a flat surface.

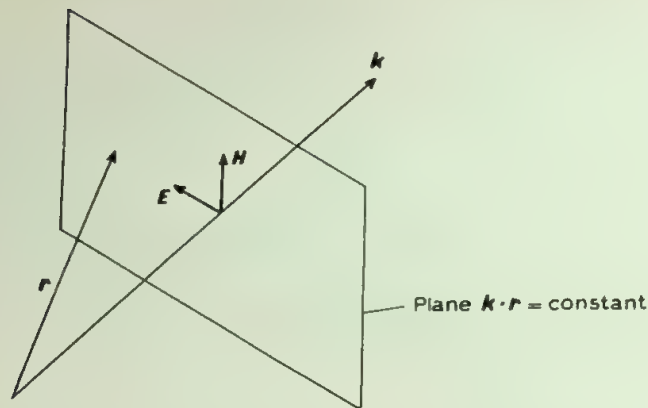


Fig. I.12

If the direction of propagation is chosen to be the x -axis, then formulae (33.5) and (33.6) can be written in the form

$$\mathbf{E} = \mathbf{E}_0 e^{i(\omega t - kx)}, \quad (33.8)$$

$$\mathbf{H} = \mathbf{H}_0 e^{i(\omega t - kx)}. \quad (33.9)$$

Formulae (33.7)–(33.9) describe plane waves propagating in the positive direction of the x -axis with a velocity

$$v = \omega/k = c. \quad (33.10)$$

Clearly at time $(t + 1)$ the phase factor at point $(x + c)$ has the same value as it had at time t at point x . Thus, v represents the velocity of propagation of the equal-phase surface and for this reason it is called the phase velocity.

The propagation of plane waves, in contrast to spherical waves, is not accompanied by a decrease in their amplitude.

We calculate the Poynting vector in a plane wave:

$$\boldsymbol{\sigma} = \frac{c}{4\pi} \mathbf{E} \times \mathbf{H} = \frac{c}{4\pi} (\mathbf{H} \times \mathbf{n}) \times \mathbf{H} = nc \frac{H^2}{4\pi} = nc \frac{E^2 + H^2}{8\pi} = nc u_0 \quad (33.11)$$

It turns out to be constant and equal to the energy flux moving with the velocity of light. The momentum of a plane electromagnetic wave is equal to

$$\mathbf{g} = \sigma/c^2 = \mathbf{n}u_0/c. \quad (33.12)$$

The expressions we have found for the field in a plane wave can also be obtained directly from the solution of the field equations in a region of space free of charges.

Since the charge density and current density in the part of space considered are equal to zero, the equations for the electromagnetic field potentials assume the following form:

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0,$$

$$\nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = 0,$$

$$\nabla \cdot \mathbf{A} + \frac{1}{c} \frac{\partial \varphi}{\partial t} = 0.$$

For simplicity we consider first the case where the field potentials depend only on the x coordinate, so that the equations for the potentials have the form

$$\frac{\partial^2 \mathbf{A}(x, t)}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}(x, t)}{\partial t^2} = 0, \quad (33.13)$$

$$\frac{\partial^2 \varphi(x, t)}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = 0, \quad (33.14)$$

$$\frac{\partial \mathbf{A}_x(x, t)}{\partial x} + \frac{1}{c} \frac{\partial \varphi}{\partial t} = 0. \quad (33.15)$$

Moreover, we assume that the field depends on time according to a simple harmonic law. We then look for the solutions of eqs. (33.13)–(33.15) in the form

$$\mathbf{A}(x, t) = \mathbf{A}_0(x) e^{i\omega t}, \quad (33.16)$$

$$\varphi(x, t) = \varphi_0(x) e^{i\omega t}. \quad (33.17)$$

The substitution of (33.16) and (33.17) into (33.13) and (33.14) gives for \mathbf{A}_0 and φ_0

$$\frac{d^2 A_0(x)}{dx^2} + \frac{\omega^2}{c^2} A_0 = 0,$$

$$\frac{d^2 \varphi_0(x)}{dx^2} + \frac{\omega^2}{c^2} \varphi_0 = 0.$$

Whence

$$A_0 = l e^{-ikx} + l' e^{+ikx},$$

$$\varphi_0 = \alpha e^{-ikx} + \alpha' e^{+ikx},$$

where $k = \omega/c$, and l, l', α, α' are constants. In this case we find for the potentials

$$A(x, t) = l e^{i(\omega t - kx)} + l' e^{i(\omega t + kx)}, \quad (33.18)$$

$$\varphi = \alpha e^{i(\omega t - kx)} + \alpha' e^{i(\omega t + kx)}. \quad (33.19)$$

The first terms in (33.18) and (33.19) represent a monochromatic plane wave propagating in the positive direction of the x -axis, while the second terms represent the same wave propagating in the negative direction of the x -axis. Which one of these waves is actually excited depends on the arrangement of the emitting system. Without restricting the general character of the treatment one can consider one of these waves, for example the first one.

The amplitudes l and α are not arbitrary, but are related by the Lorentz condition (33.15), which by substitution of (33.18) and (33.19) into it gives

$$\alpha = l_x$$

The field potentials are then written finally in the form

$$A = l e^{i(\omega t - kx)} = l e^{i\psi}, \quad (33.20)$$

$$\varphi = l_x e^{i(\omega t - kx)} = \mathbf{A} \cdot \mathbf{i} = l \cdot \mathbf{i} e^{i\psi}, \quad (33.21)$$

where \mathbf{i} is the unit vector along the x -axis, and $\psi = \omega t - kx$.

The field strengths in the plane wave have the form

$$\begin{aligned}
 \mathbf{E} &= -\frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} - \nabla \varphi = -\frac{1}{c} \frac{\partial \mathbf{A}}{\partial \psi} \frac{\partial \psi}{\partial t} - \frac{\partial \varphi}{\partial \psi} \nabla \psi = \\
 &= -\frac{\omega}{c} \dot{\mathbf{A}} - \dot{\varphi} \nabla \psi = -k \dot{\mathbf{A}} + k \dot{\varphi} \mathbf{i} = k \{ \mathbf{i} (\dot{\mathbf{A}} \cdot \mathbf{i}) - \dot{\mathbf{A}} \} = \\
 &= k (\mathbf{A} \times \mathbf{i}) \times \mathbf{i} = \mathbf{E}_0 e^{i\psi} = \mathbf{E}_0 e^{i(\omega t - kx)}, \quad (33.22)
 \end{aligned}$$

$$\mathbf{H} = \nabla \times \mathbf{A} = (\nabla \psi) \times \dot{\mathbf{A}} = k \dot{\mathbf{A}} \times \mathbf{i} = \mathbf{H}_0 e^{i(\omega t - kx)}, \quad (33.23)$$

where differentiation with respect to the argument ψ is denoted by a dot, and \mathbf{E}_0 and \mathbf{H}_0 are the amplitudes of the field strengths whose moduli are obviously $|\mathbf{E}_0| = |\mathbf{H}_0| = k|1|$. The vectors \mathbf{E}_0 and \mathbf{H}_0 are perpendicular to each other.

We write the expressions for the field components:

$$\begin{aligned}
 E_x &= 0, & H_x &= 0, \\
 E_y &= -k \dot{A}_y = H_z, & H_y &= k \dot{A}_z = -E_z, \\
 E_z &= -k \dot{A}_z = -H_y; & H_z &= -k \dot{A}_y = E_y.
 \end{aligned} \quad (33.24)$$

In these formulae, as well as in the preceding ones, the real part of the complex expressions written for the potentials and fields is implied.

In the general case, when the direction of propagation of waves does not coincide with the x -axis, the unit vector \mathbf{i} should be replaced by the unit vector in the direction of propagation \mathbf{n} (fig. 1.12), and the relations (33.22) and (33.23) are the same as (33.5) and (33.6). The values of the wave amplitudes $|\mathbf{E}_0| = |\mathbf{H}_0|$ remain completely arbitrary. They are connected with the amplitude of the waves radiated by the emitter.

It is useful to compare the expressions which we have obtained with analogous formulae obtained for another potential gauge, which is often used in the literature*.

As we have seen in § 11, field potentials allow the gauge transformation (11.1) and (11.2). We perform the transition from φ to the new value φ' ,

$$\varphi' = \varphi - \frac{1}{c} \frac{\partial \psi}{\partial t},$$

* L. D. Landau and E. M. Lifshitz, *The classical theory of fields* (Pergamon Press, London, 1962).

and choose the function $\psi = c \int \varphi dt$. Then $\varphi' = 0$, i.e. in the representation considered the field is described only by the vector potential $A'(x, t) = A + \nabla \psi$. Such a choice of ψ is not always possible but can only be made in vacuum, when $\rho = 0$ and the equation for φ allows a zero solution.

The Lorentz condition is then written in the form

$$\frac{\partial A_x}{\partial x} = 0, \quad \text{i.e.} \quad A_x = 0.$$

The solution $A_x = \text{const} \neq 0$ does not differ from $A_x = 0$, since it leads to the same values of field strengths. It is easy to see that for such a gauge we again obtain for the field components the values (33.24).

If the wave process does not have a simple periodic character, i.e. is not described by a monochromatic wave, one can easily obtain the general form of solution of the equation for the potentials φ and A . The solution is obtained in a way analogous to that in §23.

Introducing the new variables

$$\xi = x - ct, \quad \eta = x + ct,$$

we reduce the equations for the potentials to the form

$$\frac{\partial^2 A}{\partial \xi \partial \eta} = 0, \quad \frac{\partial^2 \varphi}{\partial \xi \partial \eta} = 0.$$

As the solution of the equation for the vector potential we can write

$$A = A(ct - x) + A'(ct + x),$$

where A and A' are arbitrary functions of the arguments $ct - x$ and $ct + x$, respectively. The first of these represents a general expression for a plane wave propagating with velocity c in the positive direction of the x -axis. The second one represents a wave propagating in the negative direction of the x -axis. An analogous expression is also obtained for the scalar potential.

The Lorentz condition leads to the equality

$$\varphi = A_x = A \cdot i,$$

which is valid for an arbitrary form of the functions A and φ .

Finally, it should be noted that in a vacuum one can obtain wave equations for the field vectors \mathbf{E} and \mathbf{H} directly. For this one has to take the curl of the Maxwell equation

$$\nabla \times (\nabla \times \mathbf{H}) = \nabla (\nabla \cdot \mathbf{H}) - \nabla^2 \mathbf{H} = -\nabla^2 \mathbf{H} = \frac{1}{c} \nabla \times \frac{\partial \mathbf{E}}{\partial t},$$

so that

$$\nabla^2 \mathbf{H} - \frac{1}{c^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} = 0.$$

The wave equation for \mathbf{E} can be obtained analogously.

§ 34. Plane wave polarization

Let us consider in somewhat more detail how the changes in the field vectors of a plane monochromatic wave take place. To do this we pass over from the complex to real expressions.

In formula (33.22) we write the complex amplitude \mathbf{E}_0 in the form

$$\mathbf{E}_0 = \mathbf{g}_1 + i\mathbf{g}_2,$$

where \mathbf{g}_1 and \mathbf{g}_2 are real vectors. We then have

$$\begin{aligned} \mathbf{E} &= \text{Re} \{ (\mathbf{g}_1 + i\mathbf{g}_2) [\cos(\omega t - \mathbf{k} \cdot \mathbf{r}) + i \sin(\omega t - \mathbf{k} \cdot \mathbf{r})] \} = \\ &= \text{Re} \{ [\mathbf{g}_1 \cos(\omega t - \mathbf{k} \cdot \mathbf{r}) - \mathbf{g}_2 \sin(\omega t - \mathbf{k} \cdot \mathbf{r})] + \\ &\quad + i[\mathbf{g}_1 \sin(\omega t - \mathbf{k} \cdot \mathbf{r}) + \mathbf{g}_2 \cos(\omega t - \mathbf{k} \cdot \mathbf{r})] \} = \\ &= \mathbf{g}_1 \cos(\omega t - \mathbf{k} \cdot \mathbf{r}) - \mathbf{g}_2 \sin(\omega t - \mathbf{k} \cdot \mathbf{r}). \end{aligned} \quad (34.1)$$

We now transform from the vectors \mathbf{g}_1 and \mathbf{g}_2 , orientated at an arbitrary angle to each other, to mutually perpendicular vectors \mathbf{E}_1 and \mathbf{E}_2 (fig. I.13). Let

$$\mathbf{E}_1 = \mathbf{g}_1 \cos \alpha + \mathbf{g}_2 \sin \alpha, \quad (34.2)$$

$$\mathbf{E}_2 = \mathbf{g}_1 \sin \alpha - \mathbf{g}_2 \cos \alpha, \quad (34.3)$$

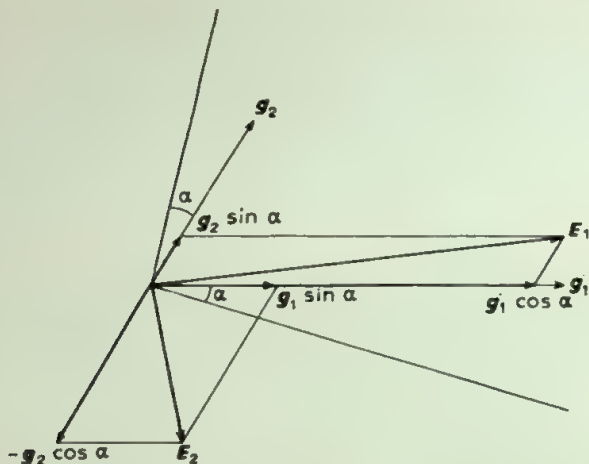


Fig. 1.13

where α is an unknown angle which is to be found. Taking the scalar product of the expressions (34.2) and (34.3), we find

$$\mathbf{E}_1 \cdot \mathbf{E}_2 = 0 = (g_1^2 - g_2^2) \sin \alpha \cos \alpha - \mathbf{g}_1 \cdot \mathbf{g}_2 (\cos^2 \alpha - \sin^2 \alpha),$$

whence

$$\tan 2\alpha = \frac{2\mathbf{g}_1 \cdot \mathbf{g}_2}{g_1^2 - g_2^2}$$

Writing

$$\mathbf{g}_1 = \mathbf{E}_1 \cos \alpha + \mathbf{E}_2 \sin \alpha,$$

$$\mathbf{g}_2 = \mathbf{E}_1 \sin \alpha - \mathbf{E}_2 \cos \alpha,$$

and substituting into (34.1), we find

$$\mathbf{E} = \mathbf{E}_1 \cos(\omega t - \mathbf{k} \cdot \mathbf{r} + \alpha) + \mathbf{E}_2 \sin(\omega t - \mathbf{k} \cdot \mathbf{r} + \alpha).$$

Let the x -axis be chosen as the direction of propagation of the waves. If the vector \mathbf{E}_1 is taken in the y -direction then the vector \mathbf{E}_2 will be in the z -direction (in the positive or negative sense). Then

$$E_y = E_1 \cos(\omega t - kx + \alpha), \quad (34.4)$$

$$E_z = E_2 \sin(\omega t - kx + \alpha). \quad (34.5)$$

The quantities E_1 and E_2 are called amplitudes, while the quantity $\psi = \omega t - kx + \alpha$ is called the phase of the wave.

From formulae (34.4) and (34.5) it is easy to eliminate the phase by writing

$$\frac{E_y^2}{E_1^2} + \frac{E_z^2}{E_2^2} = 1. \quad (34.6)$$

The expression (34.6) relates the values of the components of the vector \mathbf{E} in the plane wave. It is obvious that in a given plane $x = \text{const}$, the vector \mathbf{E} rotates in the (yz) plane in such a way that its end describes an ellipse (fig. I.14).

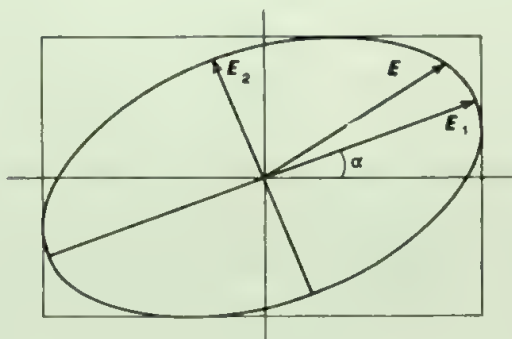


Fig. I.14

Formula (34.6) represents the equation of this ellipse. If, in particular, the amplitudes E_1 and E_2 are equal in absolute value, the vector \mathbf{E} rotates on a circle.

Since the propagation of the electromagnetic wave takes place in the direction \mathbf{n} , one can picture the change of the vector \mathbf{E} in space and time in an obvious way as the motion of its end on an elliptical (or circular) helix drawn about the line \mathbf{n} . The pitch of the helix is equal to the wavelength $\lambda = 2\pi/k$.

By virtue of eqs. (33.24) the following expressions can be written for the components of the magnetic field strength:

$$H_y = -E_2 \sin(\omega t - kx + \alpha), \quad (34.7)$$

$$H_z = E_1 \cos(\omega t - kx + \alpha). \quad (34.8)$$

Waves in which the vectors \mathbf{E} and \mathbf{H} rotate on an ellipse are called elliptically polarized, and those in which the vectors rotate on a circle are called circularly polarized.

The direction of rotation of the vector \mathbf{E} is determined by the phase α . If the rotation is clockwise for the observed looking in the direction of propagation of the wave, then such a wave is said to have left-handed polarization or positive helicity (so for example for $\alpha = \pm \frac{1}{2}\pi$ provided that \mathbf{g}_1 and \mathbf{g}_2 are directed along the positive axes).

If one of the vectors \mathbf{E}_1 or \mathbf{E}_2 is equal to zero, the change of \mathbf{E} and \mathbf{H} takes place in mutually perpendicular directions. Such waves are called plane polarized waves. According to historical tradition, the direction in which the vector \mathbf{H} oscillates is called the polarization direction. Thus, for example, a wave polarized in the z -direction has a component H_z different from zero, and a component E_y of the same magnitude, oscillating in the y -direction.

§35. Interference and the formation of wave packets

The monochromatic plane wave considered above appears to be only an idealization of real electromagnetic waves. On the one hand, the monochromatic plane wave, which is a process strictly periodic in space and time, should obviously have an infinitely large extension in space and an infinitely long duration. On the other hand, there are no strictly monochromatic emitters. As we have seen in §30, the effect of damping shows itself in the emission of frequencies differing from the natural frequency ω_0 (even if close to it). Hence to describe real wave processes it is necessary to consider the result of superposition and interference of different monochromatic plane waves.

Let us consider the superposition of an infinitely large number of monochromatic plane waves whose frequencies are continuously distributed in a narrow interval $\omega_0 - \Delta\omega \leq \omega \leq \omega_0 + \Delta\omega$, where ω_0 , called the carrier frequency, satisfies the condition $\omega_0 \gg \Delta\omega$.

We consider the amplitude of all the waves to be constant. For the strength of the electric (or magnetic) field we can write

$$\mathbf{E} = \int_{\omega_0 - \Delta\omega}^{\omega_0 + \Delta\omega} \mathbf{E}_0 e^{i(\omega t - kx)} d\omega = \mathbf{E}_0 \int_{\omega_0 - \Delta\omega}^{\omega_0 + \Delta\omega} e^{i(\omega t - kx)} d\omega.$$

For the general character of the results, which we shall need later, we assume that the wave number k is a function of the frequency ω not necessarily corresponding to the relation $k = \omega/c$ valid for electromagnetic waves in vacuum.

Setting

$$\omega = \omega_0 + (\omega - \omega_0),$$

$$k \approx k_0 + \left(\frac{dk}{d\omega}\right)_{\omega=\omega_0} (\omega - \omega_0) \equiv k_0 + \left(\frac{dk}{d\omega}\right)_0 (\omega - \omega_0),$$

we find

$$\begin{aligned} E &= E_0 \int_{\omega_0 - \Delta\omega}^{\omega_0 + \Delta\omega} e^{i(\omega_0 t - k_0 x)} \exp i(\omega - \omega_0) \left[t - \left(\frac{dk}{d\omega}\right)_0 x \right] d\omega = \\ &= E_0 e^{i(\omega_0 t - k_0 x)} \int_{-\Delta\omega}^{\Delta\omega} \exp iu \left[t - \left(\frac{dk}{d\omega}\right)_0 x \right] du = \\ &= 2E_0 \frac{\sin \Delta\omega \left[t - \left(\frac{dk}{d\omega}\right)_0 x \right]}{t - \left(\frac{dk}{d\omega}\right)_0 x} e^{i(\omega_0 t - k_0 x)} \end{aligned} \quad (35.1)$$

where the integration variable $u = \omega - \omega_0$ is introduced. We arrive in principle at the following result: the superposition of waves with a spectrum with frequencies lying in a narrow interval $2\Delta\omega$ about the carrier frequency ω_0 leads to the appearance of a wave with frequency ω_0 and wave number k_0 but with a modulated amplitude

$$A = \frac{2E_0 \sin \{ \Delta\omega / v_g (x - v_g t) \}}{v_g^{-1} (x - v_g t)}, \quad (35.2)$$

$$v_g = \left(\frac{d\omega}{dk} \right)_0. \quad (35.3)$$

The modulated amplitude has a very sharp principal maximum (fig. 1.15 where the dependence of A on $\frac{1}{2} (\Delta\omega / v_g)(x - v_g t)$ is given) at the point

$$x_m = v_g t,$$

where it is equal to

$$A_{\max} = 2\Delta\omega E_0.$$

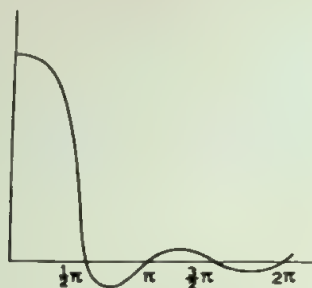


Fig. 1.15

In both directions away from the position of the maximum the value of the modulated amplitude decreases, and at points where

$$\frac{\Delta\omega}{v_g} (x - v_g t) = \pm \pi$$

the modulated amplitude reduces to zero. In addition to the principal maximum, the modulated amplitude has a number of subsidiary maxima in which, however, the amplitudes are very small in comparison with that in the principal maximum, and the heights of which decrease rapidly with increasing argument (see fig. 1.15). In practice, it can be assumed that the electromagnetic field is excited only near the principal maximum, while elsewhere in space the superposition of the waves leads to their complete mutual cancellation.

The group of waves form a wave packet. The wave packet moves in space with a velocity v_g , maintaining a limited extension in space. Hence the quantity v_g is called the group velocity of motion of the packet, as distinct from the phase velocity

$$v = \frac{\omega}{2\pi} \lambda ,$$

with which the equal-phase surface of a single monochromatic wave moves in space. It is obvious that the energy of the wave packet travels with its amplitude, i.e. with a velocity v_g . The dimensions of the wave packet are determined by the relation

$$\Delta\omega \left[t - \left(\frac{dk}{d\omega} \right)_0 x \right] \approx 2\pi . \quad (35.4)$$

For a more obvious explanation of the meaning of the equality (35.4) let us find the spatial dimensions of the packet. At a fixed moment t the field differs from zero between the points x_1 and x_2 which are a distance

$$x_1 - x_2 = \Delta x = \frac{2\pi}{\Delta\omega(dk/d\omega)_0} = \frac{2\pi}{\Delta k} \quad (35.5)$$

apart. Outside this region the field has a value close to zero.

If now a certain location $x = \text{const}$ is considered, then the duration of the time interval during which the field of the wave packet differs from zero is equal to

$$\Delta t \approx 2\pi/\Delta\omega. \quad (35.6)$$

After the lapse of time Δt the field will reduce to zero at the given location.

Thus, the wave packet has a limited spatial extension and duration, satisfying the conditions

$$\Delta t \Delta\omega \approx 2\pi, \quad \Delta x \Delta k \approx 2\pi. \quad (35.7)$$

Now the statement of the problem can be modified. Suppose we want to obtain a wave field different from zero in a certain region of space Δx . To obtain such a field from monochromatic waves it is necessary to form a wave packet.

If the dimensions of the packet are given by our condition, then according to (35.7) the superposition of monochromatic waves with wave numbers lying in an interval Δk must be carried out. The narrower the wave packet, i.e. the smaller its spatial dimensions, the larger Δk , i.e. the larger the interval of the wavelengths which must take part in the formation of the packet.

A wave packet existing for a limited time at a certain location can be considered in a completely analogous way. The shorter the required duration Δt of the packet, the larger the frequency interval $\Delta\omega$ of those monochromatic waves which must form it.

The relations found for waves propagating along the x -axis can easily be generalized to the case of waves propagating in an arbitrary direction in space. One then obtains the relations

$$\Delta x \Delta k_x \approx 2\pi, \quad \Delta y \Delta k_y \approx 2\pi, \quad \Delta z \Delta k_z \approx 2\pi, \quad \Delta t \Delta\omega \approx 2\pi. \quad (35.8)$$

The results obtained are of a great theoretical and practical significance.

A monochromatic wave completely homogeneous and infinitely extended in space and time, cannot, as was pointed out before, be realized in practice. However, by making use of an emitter with sufficiently weak damping, radiating for a sufficiently long time, one can produce in space waves whose properties are sufficiently close to those of monochromatic waves. It is clear that monochromatic waves of infinite extent cannot be used for the transmission of signals.

We understand by signals electromagnetic perturbations which in principle can be detected by means of suitable devices and which can provide information about physical events. Thus, for example, the fact of emission of radiation, beginning at a certain instant of time, is a signal. The detection of a system of charges producing the scattering of electromagnetic waves is another example of a signal.

From the above it is clear that electromagnetic waves can be used for producing signals only in the case where wave packets are formed by them. The relations (35.8) are used for the analysis of the properties required of the signal. Let, for example, a recording device need for its operation a signal whose duration is not shorter than a certain value Δt . Then the signal will be recorded only in the case where it represents a wave packet formed of monochromatic waves with frequencies distributed within an interval $\Delta\omega \geq 2\pi/\Delta t$.

We shall not dwell on other examples of the application of the relations (35.8). They play an important role in quantum mechanics.

In conclusion we stress that the expression which we obtained for the wave packet is only of an approximate character. It is valid if:

- 1) the amplitudes of all monochromatic waves forming the packet have one and the same value;
- 2) in the expansion of $k(\omega)$ in a series in $\Delta\omega$ one can restrict oneself to the first term.

The first restriction has no great significance, and finding wave packets formed by waves with different amplitudes does not present any particular difficulty.

The second requirement is fulfilled automatically for electromagnetic waves in a vacuum:

$$k = \frac{\omega}{c} = \frac{\omega_0}{c} + \frac{\Delta\omega}{c} = k_0 + \frac{1}{c} \Delta\omega.$$

However, as we shall see in quantum mechanics, in cases where the requirement 2) is not fulfilled, important consequences follow from taking into account further terms of the expansion: the form of the packet, which, in

the first approximation, is constant in time and space, turns out to be variable when further terms are taken into account. The wave packet progressively deforms and diffuses.

§ 36. Scattering of the electromagnetic waves by a free charge and by a bound electron

Let us now consider the effects arising when an electromagnetic wave is incident on a system of charged particles.

The electromagnetic field of the wave acts on the particles of the system with the Lorentz force

$$\mathbf{F} = e \left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{H}}{c} \right) \approx e\mathbf{E}.$$

Since in the electromagnetic wave $|\mathbf{H}| = |\mathbf{E}|$, the magnetic part of the Lorentz force is smaller than the electric part by the ratio v/c . Each particle of the system acquires an acceleration $\mathbf{w} = \mathbf{F}/m$, and becomes in its turn an emitter. Since each particle emits spherical electromagnetic waves, the direction of emission of these waves is not dependent upon the direction of propagation of the incident wave. Hence it is said that the initial wave is scattered. The frequencies of the emitted waves are the same as that of the incident wave. The scattering taking place without a change in the frequency is called coherent scattering. In addition to coherent scattering, in some cases there may occur scattering with a change in frequency — incoherent scattering (see Part V).

The scattering of electromagnetic waves by a bound charge can be considered as an example of the linear harmonic oscillator. Let a plane polarized, monochromatic plane wave be incident on the oscillator. The equation of motion will have the form

$$\ddot{\mathbf{r}} + \omega_0^2 \mathbf{r} = \frac{2}{3} \frac{e^2}{mc^3} \ddot{\mathbf{r}} + \frac{e\mathbf{E}_0 e^{i\omega t}}{m}$$

Assuming the damping to be weak, we can write $\ddot{\mathbf{r}} \approx -\omega_0^2 \dot{\mathbf{r}}$, so that

$$\ddot{\mathbf{r}} + \omega_0^2 \mathbf{r} + \gamma \dot{\mathbf{r}} = \frac{e}{m} \mathbf{E}_0 e^{i\omega t}, \quad (36.1)$$

where γ , as in §30, is equal to $\frac{2}{3} e^2 \omega_0^2 / mc^3$. The particular solution of eq. (36.1) which interests us is

$$\mathbf{r} = \frac{(e/m) E_0 e^{i\omega t}}{\omega_0^2 - \omega^2 + i\omega\gamma}. \quad (36.2)$$

Formula (36.2) describes the motion of the oscillator under the action of an external force.

For the power of the radiation scattered into a solid angle $d\Omega$ we find, according to formula (27.8),

$$\begin{aligned} dI &= \frac{e^2}{4\pi c^3} [\ddot{\mathbf{r}} \times \mathbf{n}]^2 d\Omega = \\ &= \frac{e^4 \omega^4 E_0^2}{4\pi m^2 c^3} \times \frac{\sin^2 \vartheta}{(\omega_0^2 - \omega^2)^2 + \omega^2 \gamma^2} \cos^2(\omega t - \delta) d\Omega. \end{aligned} \quad (36.3)$$

The scattered radiation has the same frequency as the incident radiation, but is shifted in phase by an amount

$$\delta = \arctan \frac{\gamma\omega}{\omega_0^2 - \omega^2}.$$

Introducing the incident intensity $\bar{I}_0 = cE_0^2/8\pi$ and averaging over the period $T = 2\pi/\omega$, we find

$$\overline{dI} = \frac{1}{T} \int_0^T dI dt = r_0^2 \bar{I}_0 \frac{\omega^4 \sin^2 \vartheta d\Omega}{(\omega_0^2 - \omega^2)^2 + \omega^2 \gamma^2}, \quad (36.4)$$

where r_0 is the classical electron radius (29.6). Here ϑ stands for the angle between the direction of observation \mathbf{n} and the direction of the vector of polarization \mathbf{E}_0 (fig. 1.16).

The process of scattering of electromagnetic waves is characterized by the differential cross section. By definition the differential cross section for scattering into a solid angle $d\Omega$ is the ratio of the intensity of the radiation scattered into the angle $d\Omega$ to the intensity of the incident radiation:

$$d\sigma = \frac{\overline{dI}}{\bar{I}_0}. \quad (36.5)$$

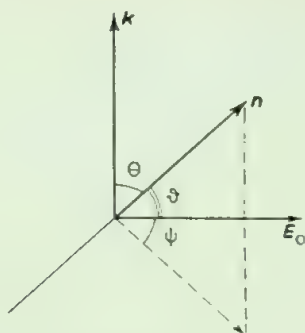


Fig. I.16

From (36.4) we find for the differential scattering cross section

$$d\sigma = \frac{r_0^2 \omega^4 \sin^2 \vartheta}{(\omega_0^2 - \omega^2)^2 + \omega^2 \gamma^2} d\Omega. \quad (36.6)$$

As is seen from formula (36.5) the differential scattering cross section has the dimensions of cm^2 (it is this which accounts for the term “cross section”). Formula (36.5) gives the cross section for the scattering of plane polarized light. The angle ϑ and the polar angles θ and ψ are mutually connected by a relation which is easily established from fig. I.16. In this drawing the polar axis z is directed along the wave vector \mathbf{k} of the incident wave, while the axis x is directed along its vector of polarization. Projecting the unit vector in the direction of observation \mathbf{n} onto the axis x , we find

$$\cos \vartheta = \sin \theta \cos \psi,$$

or

$$\sin^2 \vartheta = 1 - \sin^2 \theta \cos^2 \psi. \quad (36.7)$$

In practice it is often important to know the scattering cross section for unpolarized radiation. In order to find it, it is necessary to average the cross section (36.5) over all possible polarizations, i.e. over all possible orientations of the vector \mathbf{E}_0 in the plane (xy) . This means that it is necessary to average the expression (36.7) over all possible values of the azimuthal angle ψ , i.e. to assume

$$\sin^2 \vartheta = 1 - \sin^2 \theta \overline{\cos^2 \psi} = 1 - \frac{1}{2} \sin^2 \theta = \frac{1}{2} + \frac{1}{2} \cos^2 \theta.$$

We then obtain

$$d\sigma = \frac{r_0^2 \omega^4}{(\omega_0^2 - \omega^2)^2 + \omega^2 \gamma^2} \frac{1 + \cos^2 \theta}{2} d\Omega \quad (36.8)$$

The angular dependence in formula (36.8) shows that the strongest scattering takes place in the direction of the incident radiation ($\theta = 0$) and in the opposite direction ($\theta = \pi$).

Integrating (36.8) over the entire solid angle we obtain the total scattering cross section for unpolarized radiation

$$\sigma = \frac{8}{3} \pi \frac{r_0^2 \omega^4}{(\omega_0^2 - \omega^2)^2 + \omega^2 \gamma^2} \quad (36.9)$$

This expression is called the dispersion formula of classical electrodynamics.

The dependence of the total cross section on frequency is expressed by a curve similar to that shown in fig. I.11, with a strongly pronounced maximum for $\omega \approx \omega_0$, i.e. for the frequency of the incident radiation close to the natural frequency of the oscillator. Assuming in formula (36.9) that $\omega \approx \omega_0$, we obtain near the resonance

$$\sigma \approx \frac{8}{3} \pi \frac{r_0^2 \omega_0^2}{(\omega_0 - \omega)^2 + (\frac{1}{2} \gamma)^2} \quad (36.10)$$

The quantity γ characterizes the width of the resonance region. In particular, for exact resonance $\omega = \omega_0$, the cross section in the maximum is equal to

$$\sigma_{\max} = \frac{8}{3} \pi r_0^2 \frac{\omega_0^2}{\gamma^2} \quad (36.11)$$

Since $\gamma \ll \omega_0$, the cross section for the resonance frequency attains very large values. This phenomenon plays an important role in the optics of material media. It is called resonance fluorescence*.

The dispersion formula, obtained in the example of the scattering of radiation by a harmonic oscillator, has in reality a very general character. It is of the same form as the corresponding formula for the scattering of light by atoms obtained in quantum mechanics (Part V). We shall see that in quantum mechanics the region of applicability of the dispersion formula is not re-

* Comparing (36.10) with formula (III.4') we see that for $\gamma \rightarrow 0$ the cross section behaves as a δ -function.

stricted to the scattering of light, but also extends to a number of other systems.

Let us consider the two relations obtained from (36.9) in the limiting case of low and high frequencies. For low frequencies $\omega \ll \omega_0$

$$\sigma \approx \frac{8}{3} \pi r_0^2 \frac{\omega^4}{\omega_0^4}, \quad (36.12)$$

i.e. the cross section is proportional to the fourth power of the frequency or inversely proportional to the fourth power of the wavelength of the incident radiation.

This law of scattering is of a very general character. It is applicable when the wavelength of the light scattered is large in comparison with the dimensions of the scattering object. This case is often called Rayleigh scattering.

For high frequencies $\omega \gg \omega_0$ formula (36.9) is again simplified:

$$\sigma \approx \frac{8\pi r_0^2}{3} = \frac{8\pi e^4}{3m^2 c^4}. \quad (36.13)$$

This is called the Thomson formula.

The cross section turns out to be constant, depending neither on the frequency of the radiation scattered nor on the properties of the oscillator. This fact has a simple meaning. At high frequencies the force acting on the charge due to the field is very large in comparison with the quasi-elastic force. The electron scatters as a free particle. According to the Thomson formula, the cross section σ is a universal constant determined by the classical radius of the electron. It is clear that the Thomson formula describes the scattering by electrons belonging to any system, for example atoms, if one can disregard the forces binding the electrons in the atoms and consider the electrons to be free. The scattering of radiation by heavy nuclei can be disregarded, since the cross section is inversely proportional to the square of the mass of the scatterer.

The universal character of formula (36.13) makes it one of the fundamental relations of classical electrodynamics. It has been put to a thorough experimental test, the results of which are shown in fig. 1.17.

We see that the ratio $\sigma_{\text{exp}}/\sigma_{\text{Thom}} = 1$ only for wavelengths larger than about 2 \AA .

For smaller wavelengths the classical treatment of scattering processes turns out to be inapplicable. Here we encounter the fact mentioned in §28. Although the limit of applicability of classical electrodynamics, inherent in

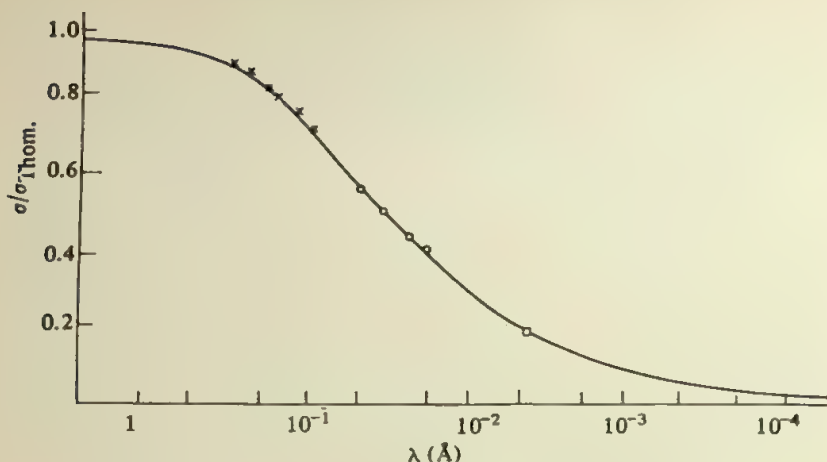


Fig. 1.17

the theory, occurs at the scale of wavelengths $\lambda \approx r_0 \approx 10^{-5}$ Å, actually it already appears at a scale 10^5 times larger. This is associated, as we have already stressed, with the manifestation of quantum effects. The trend of the cross section in fig. 1.17 is in excellent agreement with the predictions of the quantum theory of radiation (Part V, the Klein-Nishina formula).

§37. Absorption of radiation

In addition to scattering, the radiation interacting with matter undergoes absorption. In classical electrodynamics this effect can be calculated for the oscillator model. We shall restrict ourselves to the case where the frequency of the incident radiation is close to the resonance frequency, when the absorption is greatest. However, the frequency of the radiation cannot then be assumed to be exactly equal to the natural frequency of the oscillator. From what follows it can be seen that in such a case the calculation would not lead to a well defined result. Hence it should be assumed that the incident radiation has a continuous frequency distribution.

We can expand the incident field in a Fourier integral, writing

$$\mathbf{E}(t) = \int \mathbf{E}(\omega) e^{i\omega t} d\omega.$$

Substituting this into the equation of motion of the oscillator (36.1) and expanding the displacement of the latter in a Fourier integral,

$$\mathbf{r}(t) = \int \mathbf{r}(\omega) e^{i\omega t} d\omega,$$

it is easy to find that

$$\mathbf{r}(\omega) = \frac{e}{m} \frac{\mathbf{E}(\omega)}{\omega_0^2 - \omega^2 + i\omega\gamma}. \quad (37.1)$$

The energy loss is obviously equal to the total work done by the field on the oscillator. The latter can be calculated from the formula

$$-\Delta E = W = \int \mathbf{F} \cdot \mathbf{v} dt = e \int \mathbf{E}(t) \cdot \dot{\mathbf{r}}(t) dt. \quad (37.2)$$

Proceeding in an analogous way as in deriving the Parseval formula (II.9) we find, using $\mathbf{E}^*(\omega) = \mathbf{E}(-\omega)$, $\mathbf{r}^*(\omega) = \mathbf{r}(-\omega)$ and eq. (37.1):

$$\begin{aligned} -\Delta E &= \frac{1}{2} e \left[\int_{-\infty}^{+\infty} dt \mathbf{E}(t) \cdot \int_{-\infty}^{+\infty} i\omega \mathbf{r}(\omega) e^{i\omega t} d\omega + \int_{-\infty}^{+\infty} dt \dot{\mathbf{r}}(t) \cdot \int_{-\infty}^{+\infty} \mathbf{E}(\omega) e^{i\omega t} d\omega \right] \\ &= \frac{2\pi e}{2} \int_{-\infty}^{+\infty} i\omega [\mathbf{E}^*(\omega) \cdot \mathbf{r}(\omega) - \mathbf{E}(\omega) \cdot \mathbf{r}^*(\omega)] d\omega = \\ &= \frac{4\pi e^2}{m} \int_0^{\infty} \frac{\omega^2 \gamma |\mathbf{E}(\omega)|^2 d\omega}{(\omega_0^2 - \omega^2)^2 + \omega^2 \gamma^2} \approx \frac{\pi e^2}{m} \int_0^{\infty} \frac{\gamma |\mathbf{E}(\omega)|^2 d\omega}{(\omega_0 - \omega)^2 + (\frac{1}{2}\gamma)^2}. \end{aligned} \quad (37.3)$$

The integrand has a sharp maximum in the region of resonance. We assume that the spectral distribution of the radiation absorbed in the region of resonance varies slowly in comparison with the resonance factor. Then the integral can be calculated approximately. Substituting $\omega = \omega_0 + \frac{1}{2}\gamma x$ we get

$$\begin{aligned} \int_0^{\infty} \frac{\gamma |\mathbf{E}(\omega)|^2 d\omega}{(\omega_0 - \omega)^2 + (\frac{1}{2}\gamma)^2} &= \int_{-2\omega_0/\gamma}^{\infty} \frac{\frac{1}{2}\gamma^2 |\mathbf{E}(\omega_0 + \frac{1}{2}\gamma x)|^2 dx}{(\frac{1}{2}\gamma x)^2 + (\frac{1}{2}\gamma)^2} \approx \\ &\approx 2 \int_{-\infty}^{+\infty} |\mathbf{E}(\omega_0 + \frac{1}{2}\gamma x)|^2 \frac{dx}{x^2 + 1} \approx 2\pi |\mathbf{E}(\omega_0)|^2 \end{aligned} \quad (37.4)$$

We have replaced the lower limit of the integration by infinity, since $\gamma \ll \omega_0$. Substituting (37.4) into (37.3), we find

$$-\Delta E = \frac{2\pi^2 e^2}{m} |E(\omega_0)|^2 = 2\pi^2 c^2 r_0 |E(\omega_0)|^2. \quad (37.5)$$

The energy absorbed turns out to be independent of the physical properties of the absorbing system, and depends only on the position of the resonance frequency ω_0 . Hence the expression obtained is of a very general character as in the case of the dispersion formula. Very similar expressions for absorption are also obtained in the quantum theory of radiation (see Part V).

§38*. Canonical form of the field equations

In passing over to the quantum theory of the electromagnetic field it is convenient to give to the electromagnetic field equations a form very similar to that of the equations of mechanics. It turns out that the electromagnetic field can be compared with a certain mechanical system, and that D'Alembert's equations can take the form of Hamilton's equations describing the motion of this mechanical system. We shall confine ourselves to the case of the electromagnetic field in vacuum.

Making use of the Coulomb potential gauge described in §11, for which $\varphi = 0$, we write the field equations in the form

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0, \quad (38.1)$$

$$\nabla \cdot \mathbf{A} = 0. \quad (38.2)$$

The region in which these field equations are valid is the space which is free of charges and currents. At the limit of this region a boundary condition must be given. Such a condition may take, for example, the expression for \mathbf{A} in the form of a retarded potential. However, we shall not be interested in the amplitude of electromagnetic waves, and shall confine ourselves to the consideration of the field in vacuum at a large distance from currents and charges. We divide the entire region of space which is free of charge into a set of cubes (we shall call them the normalization cubes) of edge L , and consider the field inside one of the cubes. Then the boundary condition must be given at the surface of the normalization cube.

As we know, the electromagnetic field in a vacuum represents a set of travelling waves. The solution in the form of travelling waves is obtained if it is assumed that the vector potential \mathbf{A} and its derivatives have equal values at opposite faces of the cube. This is equivalent to the requirement: \mathbf{A} is a periodic function of the variables x, y, z with a period L :

$$\mathbf{A}(x, y, z) = \mathbf{A}(x + L, y + L, z + L). \quad (38.3)$$

Having obtained the solution of eqs. (38.1) and (38.2) in a normalization cube with edge L with the boundary condition (38.3), we can write the solution in the entire space by simply repeating the solution in the initial normalization cube. The final results will not depend on the choice of L .

We look for the solution of (38.1) and (38.2) in a normalization cube in the form of a set of expressions of the form $q_i(t) \mathbf{A}_i(\mathbf{r})$, where $q_i(t)$ depends only on time, and $\mathbf{A}_i(\mathbf{r})$ depends only on coordinates.

Each of the vector functions $\mathbf{A}_i(\mathbf{r})$ represents a wave in the normalization cube. Since the volume of the latter is finite, it can contain a denumerably great number of standing or travelling waves. Thus, we set

$$\mathbf{A}(\mathbf{r}, t) = \sum_i q_i(t) \mathbf{A}_i(\mathbf{r}), \quad (38.4)$$

where the index i runs over an infinite but discrete number of values. This means that the vector potential of the field in the normalization cube can be expanded in a Fourier series. The number of terms in the sum (38.4), i.e. the number of waves in the normalization cube, is infinitely large.

Substituting (38.4) into (38.1), we find

$$\sum_i \left(q_i(t) \nabla^2 \mathbf{A}_i(\mathbf{r}) - \frac{1}{c^2} \ddot{q}_i(t) \mathbf{A}_i(\mathbf{r}) \right) = 0. \quad (38.5)$$

Since all waves forming the superposition (38.4) are independent, the equality (38.5) must hold for each of the waves, i.e.

$$q_i(t) \nabla^2 \mathbf{A}_i(\mathbf{r}) - \frac{1}{c^2} \ddot{q}_i(t) \mathbf{A}_i(\mathbf{r}) = 0. \quad (38.6)$$

We rewrite (38.6) in the form

$$\frac{c^2 \nabla^2 [\mathbf{A}_i(\mathbf{r})]_k}{[\mathbf{A}_i(\mathbf{r})]_k} = \frac{\ddot{q}_i(t)}{q_i(t)}, \quad k = x, y, z. \quad (38.7)$$

Since $[A_i(\mathbf{r})]_k$ and $q_i(t)$ are functions of different variables, the equality (38.7) can hold only if its right-hand side and left-hand side are separately equal to a constant, called the separation constant. For reasons which will be clear from what follows, this constant must be an essentially negative quantity. Denoting it by $(-\omega_i^2)$, we find

$$\nabla^2 A_i(\mathbf{r}) + \frac{\omega_i^2}{c^2} A_i(\mathbf{r}) = 0 \quad (38.8)$$

$$\ddot{q}_i(t) + \omega_i^2 q_i(t) = 0. \quad (38.9)$$

Expressions of the type

$$A_i \sim \mathbf{e}_j \sin \mathbf{k}_i \cdot \mathbf{r}, \quad (38.10)$$

$$A_i \sim \mathbf{e}_j \cos \mathbf{k}_i \cdot \mathbf{r},$$

serve as the solutions of eq. (38.8) satisfying the conditions of periodicity (38.3).

The components of the vector \mathbf{k}_i must assume a discrete number of values

$$k_{ix} = \frac{2\pi n_{i1}}{L}, \quad k_{iy} = \frac{2\pi n_{i2}}{L}, \quad k_{iz} = \frac{2\pi n_{i3}}{L}, \quad (38.11)$$

where n_{i1} , n_{i2} and n_{i3} are positive integers. The whole set of such values of the components of \mathbf{k}_i ensures the presence of nodes or antinodes at the faces of the normalization cube. The absolute value of the vector \mathbf{k} is by definition equal to $|\mathbf{k}_i| = \omega_i/c$. The vector \mathbf{e}_j is the unit polarization vector, which can assume two values ($j = 1, 2$) for a given \mathbf{k}_i .

Instead of sines and cosines in formula (38.10) we could take their arbitrary linear combination.

Eq. (38.2) leads to the requirement

$$\mathbf{e}_j \cdot \mathbf{k}_i = 0, \quad (38.12)$$

meaning (see §33) that the waves are transverse.

The whole set of functions A_i is given completely by eqs. (38.1) and (38.2) and the boundary condition (38.3). It appears to be the same for any fields in the normalization cube.

For the actual determination of the vector potential at a given point of

space at a definite instant of time it is necessary to give the set of all time amplitudes $q_i(t)$. This means that the state of the field is characterized by defining an infinite set of amplitudes $q_i(t)$.

The latter are determined by eq. (38.9), which is the same as the equation of motion of the linear harmonic oscillator. The quantity ω_i represents the frequency of this oscillator. The value of $q_i(t)$ determines the state of the i th oscillator at any moment t . The definition of the whole set of the variable amplitudes $q_i(t)$ is equivalent to the definition of the state of a whole set of an infinitely large number of oscillators with frequencies ω_i . If the state of all the oscillators at given moment t is known, then the field is also known at this moment. Thus, the electromagnetic field can formally be replaced by a mechanical system with an infinitely large number of degrees of freedom — a set of an infinitely large number of oscillators usually called field oscillators. The whole set of states of the field oscillators $q_i(t)$ characterizes the state of this mechanical system and, at the same time, the state of the field.

The quantities $q_i(t)$ can be considered as the set of coordinates of a mechanical system whose equations of motion can be written in the form of Hamilton's equations. One can ascribe to this system the Hamiltonian function (which is the same as its energy):

$$H = \frac{1}{2} \sum (p_i^2 + \omega_i^2 q_i^2), \quad (38.13)$$

where p_i is the momentum conjugate to the coordinate q_i . The mass of all field oscillators is assumed to be equal to unity.

Indeed, Hamilton's equations for the i th degree of freedom (i.e. the i th oscillator) read:

$$\dot{q}_i = \partial H / \partial p_i = p_i, \quad \dot{p}_i = -\partial H / \partial q_i = -\omega_i^2 q_i,$$

whence the eq. (38.9) for the quantity q_i follows directly.

A particular solution of the field eqs. (38.1) and (38.2) can be written in the form

$$A_{ij} \sim C_{1i} \mathbf{e}_j \cos \omega_i t \sin \mathbf{k}_i \cdot \mathbf{r} + C_{2i} \mathbf{e}_j \sin \omega_i t \cos \mathbf{k}_i \cdot \mathbf{r}. \quad (38.14)$$

If one chooses properly the normalization of the linear combination of solutions of the type (38.14), from which the general solution (38.4) of the field equations (38.1) and (38.2) is formed, one can make the energy of the field in the normalization cube equal to the energy of the system of field oscillators (38.13). For such a normalization of the vector potential we can

assume that the system of field oscillators is completely equivalent to the electromagnetic field. The state of the set of oscillators is uniquely related to the state of the field.

The necessary linear combination of solutions of the type (38.14) has the following form:

$$\begin{aligned} A &= \sqrt{\frac{4\pi}{L^3}} \sum_i \sum_{j=1}^2 \mathbf{e}_j \frac{1}{k_i} (-\dot{q}_i(t) \sin \mathbf{k}_i \cdot \mathbf{r} + \omega_i q_i \cos \mathbf{k}_i \cdot \mathbf{r}) = \\ &= \sqrt{\frac{4\pi}{L^3}} \sum_i \sum_{j=1}^2 \mathbf{e}_j \frac{1}{k_i} (-p_i \sin \mathbf{k}_i \cdot \mathbf{r} + \omega_i q_i \cos \mathbf{k}_i \cdot \mathbf{r}). \end{aligned} \quad (38.15)$$

Indeed, we shall verify the fact that the energy of the field in the normalization cube satisfies the requirement

$$E = \int \frac{\mathbf{E}^2 + \mathbf{H}^2}{8\pi} dV = \frac{1}{2} \sum_i \sum_{j=1}^2 (p_i^2 + \omega_i^2 q_i^2) \quad (38.16)$$

Oscillators with different polarizations are considered to be different. The total number of field oscillators is obtained by summing over i and j . This amounts to doubling the result.

From (38.15) it follows that

$$\begin{aligned} \mathbf{E} &= -\frac{1}{c} \frac{\partial A}{\partial t} = \\ &= -\frac{1}{c} \sqrt{\frac{4\pi}{L^3}} \sum_i \sum_{j=1}^2 \mathbf{e}_j \frac{1}{k_i} (-\ddot{q}_i \sin \mathbf{k}_i \cdot \mathbf{r} + \omega_i \dot{q}_i \cos \mathbf{k}_i \cdot \mathbf{r}) = \\ &= -\sqrt{\frac{4\pi}{L^3}} \sum_i \sum_j \mathbf{e}_j \frac{1}{ck_i} (\omega_i^2 q_i \sin \mathbf{k}_i \cdot \mathbf{r} + \omega_i p_i \cos \mathbf{k}_i \cdot \mathbf{r}) = \\ &= -\sqrt{\frac{4\pi}{L^3}} \sum_i \sum_j \mathbf{e}_j (\omega_i q_i \sin \mathbf{k}_i \cdot \mathbf{r} + p_i \cos \mathbf{k}_i \cdot \mathbf{r}) = \\ &= \sum_i \mathbf{E}_i, \end{aligned} \quad (38.17)$$

$$\begin{aligned}
\mathbf{H} &= \nabla \times \mathbf{A} = \sqrt{\frac{4\pi}{L^3}} \sum_i \sum_{j=1}^2 \frac{1}{k_i} \{ -q_i \nabla \times (\mathbf{e}_j \sin \mathbf{k}_i \cdot \mathbf{r}) + \\
&\quad + \omega_i q_i \nabla \times (\mathbf{e}_j \cos \mathbf{k}_i \cdot \mathbf{r}) \} = \\
&= -\sqrt{\frac{4\pi}{L^3}} \sum_i \sum_j \frac{1}{k_i} (p_i \mathbf{k}_i \times \mathbf{e}_j \cos \mathbf{k}_i \cdot \mathbf{r} + \omega_i q_i \mathbf{k}_i \times \mathbf{e}_j \sin \mathbf{k}_i \cdot \mathbf{r}) = \\
&= -\sqrt{\frac{4\pi}{L^3}} \sum_i \sum_j \frac{\mathbf{k}_i \times \mathbf{e}_j}{k_i} (p_i \cos \mathbf{k}_i \cdot \mathbf{r} + \omega_i q_i \sin \mathbf{k}_i \cdot \mathbf{r}) = \\
&= \sum_i \frac{\mathbf{k}_i \times \mathbf{E}_i}{k_i}. \tag{38.18}
\end{aligned}$$

Let us work out the expression for the integrals $(1/8\pi) \int \mathbf{E}^2 dV$ and $(1/8\pi) \int \mathbf{H}^2 dV$. From (38.17) it follows that

$$\frac{1}{8\pi} \int \mathbf{E}^2 dV = \frac{1}{2L^3} \int \left\{ \sum_i \sum_j \mathbf{e}_j (\omega_i q_i \sin \mathbf{k}_i \cdot \mathbf{r} + p_i \cos \mathbf{k}_i \cdot \mathbf{r}) \right\}^2 dV$$

In calculating the integrals we make use of the obvious relations

$$\begin{aligned}
\int \sin \mathbf{k}_i \cdot \mathbf{r} \sin \mathbf{k}_{i'} \cdot \mathbf{r} dV &= \int \sin \mathbf{k}_i \cdot \mathbf{r} \cos \mathbf{k}_{i'} \cdot \mathbf{r} dV = \\
&= \int \cos \mathbf{k}_i \cdot \mathbf{r} \cos \mathbf{k}_{i'} \cdot \mathbf{r} dV = \\
&= \int \sin \mathbf{k}_i \cdot \mathbf{r} \cos \mathbf{k}_{i'} \cdot \mathbf{r} dV = 0. \\
\int \sin^2 \mathbf{k}_i \cdot \mathbf{r} dV &= \int \cos^2 \mathbf{k}_i \cdot \mathbf{r} dV = \frac{1}{2} V = \frac{1}{2} L^3.
\end{aligned}$$

Hence it is easy to obtain the expression

$$\frac{1}{8\pi} \int \mathbf{E}^2 dV = \frac{1}{4} \sum_i \sum_j (p_i^2 + \omega_i^2 q_i^2).$$

Quite analogously one obtains from (38.18):

$$\begin{aligned} \frac{1}{8\pi} \int H^2 dV &= \frac{1}{2L^3} \int \left\{ \sum_i \sum_j \mathbf{e}_j \frac{1}{k_i} (p_i \mathbf{k}_i \times \mathbf{e}_j \cos \mathbf{k}_i \cdot \mathbf{r} + \right. \\ &\quad \left. + \omega_i q_i \mathbf{k}_i \times \mathbf{e}_j \sin \mathbf{k}_i \cdot \mathbf{r}) \right\}^2 dV = \\ &= \frac{1}{4} \sum_i \sum_j (p_i^2 + \omega_i^2 q_i^2). \end{aligned}$$

Adding the two expressions, we arrive at the equality (38.16).

For the normalization method which we have chosen — i.e. the selection of the factors before $\sin \mathbf{k}_i \cdot \mathbf{r}$ and $\cos \mathbf{k}_i \cdot \mathbf{r}$ — the field in the normalization volume can be compared with a set of field oscillators whose energy is the same as that of the field. The state of the oscillators — the set of the quantities $q_i(t)$ at a moment t — determines the value of the vector potential $\mathbf{A}(\mathbf{r}, t)$.

Thus, the electromagnetic field in a finite volume is formally equivalent to a mechanical system with an infinitely large but denumerable number of degrees of freedom — a set of field oscillators — while the field equations are formally equivalent to Hamilton's equations of motion of the field oscillators. The Hamiltonian function of the equivalent system of oscillators is often called simply the Hamiltonian field function, while the expansion of the vector potential (38.15) is called the expansion of the field into oscillators. It should be stressed that within the framework of classical electrodynamics the expansion of the field into oscillators is a mathematical device. Field oscillators cannot be associated with the oscillations of any real particles. However, this expansion plays a most important role in the quantum theory of the electromagnetic field (see Part V).

In the quantum theory of the electromagnetic field the expansion (38.15) is often expressed in terms of exponential functions. Later, in Part V, we shall need such a representation. We write the expression for the vector potential $\mathbf{A}(\mathbf{r}, t)$ in the form

$$\mathbf{A}(\mathbf{r}, t) = \sum_{\lambda} (b_{\lambda} \mathbf{A}_{\lambda} + b_{\lambda}^* \mathbf{A}_{\lambda}^*). \quad (38.19)$$

Here the following notations are introduced:

$$\begin{aligned} \mathbf{A}_{\lambda} &= \mathbf{e}_{\lambda} \sqrt{\frac{4\pi c^2}{L^3}} e^{i\mathbf{k}_{\lambda} \cdot \mathbf{r}}, \\ \mathbf{A}_{\lambda}^* &= \mathbf{e}_{\lambda} \sqrt{\frac{4\pi c^2}{L^3}} e^{-i\mathbf{k}_{\lambda} \cdot \mathbf{r}}. \end{aligned}$$

The index λ , over which the summation in (38.19) is carried out, replaces the indices i and j in (38.15), i.e. it runs over a double (in comparison with i) series of values, corresponding to the two directions of the polarization vector (\mathbf{e}_1 and \mathbf{e}_2). The term $b_\lambda \mathbf{A}_\lambda$ characterizes the wave propagating (travelling) in the positive direction of the vector \mathbf{k}_λ . The second term $b_\lambda^* \mathbf{A}_\lambda^*$ represents the wave propagating in the opposite direction ($-\mathbf{k}_\lambda$). Thus, the waves having the wave vectors (\mathbf{k}_λ) and ($-\mathbf{k}_\lambda$) are considered to be different waves. The components of the vectors (\mathbf{k}_λ) and ($-\mathbf{k}_\lambda$) assume the values given by (38.11), but with both positive and negative values of the integers n_{i1} , n_{i2} , n_{i3} . This means that to one standing wave there correspond two travelling waves — in the positive and negative directions.

Substituting \mathbf{A}_λ and \mathbf{A}_λ^* into (38.19), we rewrite this formula in the form

$$\mathbf{A}(\mathbf{r}, t) = \sqrt{\frac{4\pi c^2}{L^3}} \sum_{\lambda} \mathbf{e}_{\lambda} (b_{\lambda} e^{i\mathbf{k}_{\lambda} \cdot \mathbf{r}} + b_{\lambda}^* e^{-i\mathbf{k}_{\lambda} \cdot \mathbf{r}}).$$

Comparing the above expression with (38.15) and equating coefficients of $e^{i\mathbf{k}_{\lambda} \cdot \mathbf{r}}$ and $e^{-i\mathbf{k}_{\lambda} \cdot \mathbf{r}}$, we find

$$\begin{aligned} b_{\lambda} &= \frac{1}{2\omega_{\lambda}} (ip_{\lambda} + \omega_{\lambda} q_{\lambda}), \\ b_{\lambda}^* &= \frac{1}{2\omega_{\lambda}} (-ip_{\lambda} + \omega_{\lambda} q_{\lambda}). \end{aligned} \tag{38.20}$$

In this notation the field energy is written in the form

$$E = \frac{1}{2} \sum_i \sum_j (p_i^2 + \omega_i^2 q_i^2) = \frac{1}{2} \sum_{\lambda} (p_{\lambda}^2 + \omega_{\lambda}^2 q_{\lambda}^2) = 2 \sum_{\lambda} b_{\lambda} b_{\lambda}^* \omega_{\lambda}^2. \tag{38.21}$$

Later we shall also need to calculate the number of field oscillators with given frequency and given polarization. It is obviously equal to the number of travelling waves in the volume $V = L^3$. In order to find the number of oscillators it is convenient to make use of a simple geometrical construction. We choose n_1 , n_2 and n_3 in formula (38.11) to be the coordinates in an imaginary space of the numbers (n_1, n_2, n_3) , omitting the suffixes i for clarity.

Fig. 1.18 presents a part of this space. To each possible value of n_1 and n_2 in this drawing there corresponds a point. We introduce the quantity

$$n = \sqrt{n_1^2 + n_2^2 + n_3^2}.$$

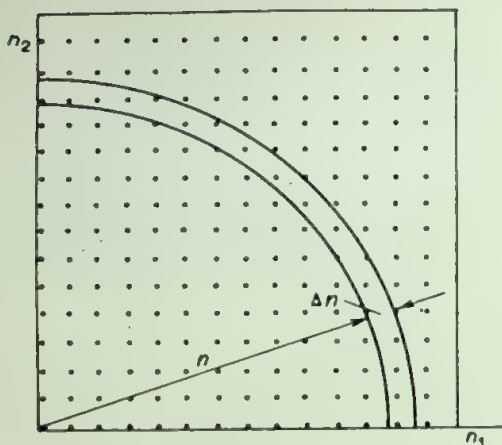


Fig. 1.18

If the numbers n_1, n_2, n_3 are sufficiently large, then the points representing them lie very close to each other and fill the entire space in an almost continuous way. The quantity n , as a function of n_1, n_2, n_3 , will vary almost continuously, and in fig. 1.18 will be represented by a radius vector.

Since to each set of three numbers n_1, n_2, n_3 there corresponds a definite value of k given by

$$k = |\mathbf{k}| = \frac{2\pi}{L} \sqrt{n_1^2 + n_2^2 + n_3^2} = n \frac{2\pi}{L},$$

the number of waves with a k lying in the interval between k and $k + dk$ is equal to the number of the values of n lying in the interval between n and $n + dn$. The latter is equal to the number of representative points within a spherical layer lying between spherical surfaces with radii $n, n + dn$. For this number we have, obviously, the value

$$g(n) dn = 4\pi n^2 dn.$$

Thus, the number of travelling waves or the number of field oscillators with $|\mathbf{k}|$ lying in the interval $k, k + dk$ and with a given polarization in a volume $V = L^3$ is equal to

$$g(k) dk = 4\pi n^2 dn = \frac{4\pi k^2 dk}{(2\pi)^3} L^3$$

The number of oscillators with a given polarization and a frequency lying in the interval $\omega, \omega + d\omega$ is

$$g(\omega) d\omega = \frac{4\pi\omega^2 d\omega}{(2\pi c)^3} L^3 \quad (38.22)$$

Hence one has to make use of the formula for the number of field oscillators with given frequency, given polarization and a direction of the vector \mathbf{k} lying within a solid angle $d\Omega$. The number of such oscillators is equal to

$$g(\omega) d\omega \frac{d\Omega}{4\pi} = \frac{\omega^2 d\omega L^3 d\Omega}{(2\pi c)^3}. \quad (38.23)$$

We shall need the expression for the momentum of radiation in a volume L^3 . Writing it on the basis of (13.11), (38.17) and (38.18) in the form

$$\begin{aligned} \mathbf{G} &= \frac{1}{4\pi c} \int \mathbf{E} \times \mathbf{H} dV = \frac{1}{4\pi c} \int \left[\sum_{\lambda} \mathbf{E}_{\lambda} \times \sum_{\mu} \frac{\mathbf{k}_{\mu} \times \mathbf{E}_{\mu}}{k_{\mu}} \right] dV = \\ &= \frac{1}{4\pi c} \sum_{\lambda} \frac{\mathbf{k}_{\lambda}}{k_{\lambda}} \int E_{\lambda}^2 dV = \int \sum_{\lambda} \frac{\mathbf{k}_{\lambda} E_{\lambda}^2}{4\pi c k_{\lambda}} dV = \sum_{\lambda} \mathbf{g}_{\lambda}, \end{aligned}$$

we can ascribe to each oscillator an energy

$$\epsilon_{\lambda} = 2b_{\lambda} b_{\lambda}^* \omega_{\lambda}^2 = p_{\lambda}^2 + \omega_{\lambda}^2 q_{\lambda}^2, \quad (38.24)$$

and a momentum

$$\mathbf{p}_{\lambda} = \mathbf{k}_{\lambda} \epsilon_{\lambda} / c k_{\lambda} = \mathbf{g}_{\lambda}, \quad (38.25)$$

so that

$$E = \sum \epsilon_{\lambda}, \quad \mathbf{G} = \sum \mathbf{p}_{\lambda}.$$

The Motion of Particles in Electromagnetic Fields

§39. The motion of charged particles in constant electric and magnetic fields

One of the branches of electrodynamics which is important from the practical point of view is the theory of motion of charged particles in electric and magnetic fields. The theory of the motion of charged particles in electromagnetic fields is at the basis of electronics, accelerator technology, electron and proton microscopy, mass-spectrography, investigation of reactions in plasma, and experimental facilities for the investigation of thermonuclear phenomena. It is very important for a number of other fields of physics — astrophysics, physics of cosmic rays and so on.

In this book we shall restrict ourselves to the consideration of the simplest problems. We shall assume that the field in which a particle is placed has a strength which is very large compared with that of the particle itself. In other words, we consider the particle whose motion interests us as a test particle which does not distort the given external field.

We begin with the motion of a charged particle in a uniform electric field constant in time. The equations of motion have the form

$$m \frac{d^2 \mathbf{r}}{dt^2} = e \mathbf{E}.$$

If at the initial instant $t = 0$ the charged particle was at rest, then, taking the direction of the field along the x -axis we have

$$m\ddot{x} = eE,$$

$$m\ddot{y} = 0;$$

it can be shown that

$$v \equiv \dot{x} = \pm \sqrt{2eV/m + v_0^2} = \pm \sqrt{2eV/m}, \quad (39.1)$$

where $V = (\varphi_2 - \varphi_1)$ is the accelerating (or decelerating) potential difference traversed by the particle. Here it is assumed that $v_0 = 0$.

If the particle at the initial instant was at the origin and had a velocity v_0 directed at an angle θ to the y -axis, then the double integration of the equations of motion gives:

$$\dot{x} = \frac{etE}{m} + v_0 \sin \theta,$$

$$x = \frac{et^2E}{2m} + (v_0 \sin \theta)t,$$

$$\dot{y} = v_0 \cos \theta,$$

$$y = (v_0 \cos \theta)t.$$

Eliminating t from the expressions written for y and x , we find the equation of the trajectory:

$$x = (\tan \theta)y + \frac{eEy^2}{2m(v_0 \cos \theta)^2}. \quad (39.2)$$

As was to be expected, the particle moves in a parabola.

Let us now consider the motion of a particle in a constant uniform magnetic field. We orientate the z -axis along the direction of the field. The equations of motion have the form

$$m\ddot{\mathbf{r}} = \frac{e}{c} \mathbf{v} \times \mathbf{H}, \quad (39.3)$$

or, taking the components,

$$\ddot{x} = \frac{e}{mc} yH, \quad (39.4)$$

$$\ddot{y} = -\frac{e}{mc} xH, \quad (39.5)$$

$$\ddot{z} = 0. \quad (39.6)$$

Eq. (39.6) means that the magnetic field directed along the z -axis has no effect on the component of the motion of a particle in this direction. We look for the solution of eqs. (39.4) and (39.5) in the form

$$\dot{x} = A \cos(\omega_C t + \alpha),$$

$$y = B \sin(\omega_C t + \alpha).$$

We then have

$$-\omega_C A = \frac{eH}{mc} B,$$

$$-\frac{eH}{mc} A = \omega_C B,$$

from which we find

$$\omega_C = eH/mc, \quad A = -B. \quad (39.7)$$

Thus, we can write

$$\begin{aligned}\dot{x} &= A \cos(\omega_C t + \alpha) = A \cos\left(\frac{eH}{mc} t + \alpha\right), \\ \dot{y} &= -A \sin(\omega_C t + \alpha) = -A \sin\left(\frac{eH}{mc} t + \alpha\right).\end{aligned}\tag{39.8}$$

It is obvious that

$$A^2 = v_x^2 + v_y^2 = v_{\perp}^{(0)2},$$

where $v_{\perp}^{(0)}$ is the initial velocity in the plane (xy).

Integrating the expressions obtained once more, we have

$$\begin{aligned}x &= x_0 + \frac{v_{\perp}^{(0)}}{\omega_C} \sin(\omega_C t + \alpha), \\ y &= y_0 + \frac{v_{\perp}^{(0)}}{\omega_C} \cos(\omega_C t + \alpha).\end{aligned}$$

Eliminating the time from the above relations, we find that the particle moves in a circle

$$(x - x_0)^2 + (y - y_0)^2 = \frac{v_{\perp}^{(0)2}}{\omega_C^2} = R_C^2.$$

The frequency of rotation of the particle, given by formula (39.7), is called the cyclotron frequency. The cyclotron frequency, equal to double the Larmor frequency, does not depend on the initial velocity of the particle and is determined by the ratio e/m . The radius of the circle

$$R_C = \frac{v_{\perp}^{(0)}}{\omega_C} = \frac{mc v_{\perp}^{(0)}}{eH}\tag{39.9}$$

has the following simple meaning: on a circle of radius R_C the centrifugal force and the Lorentz force are balanced,

$$\frac{m(v_{\perp}^{(0)})^2}{R_C} = \frac{e}{c} v_{\perp}^{(0)} H.$$

If at the initial moment the particle had a velocity component $v_z^{(0)}$ along

the z -axis in addition to its velocity in the plane (xy), then it would move uniformly along the direction of the magnetic field. The superposition of the two motions, the uniform one along the z -axis and the rotation in the plane (xy), leads to a helical trajectory of the particle in the longitudinal field. The turns of the helix lie on a cylindrical surface of radius R_C with its axis parallel to the z -axis.

For the motion in a constant magnetic field \mathbf{H} the following conservation laws hold:

1. The total energy of the particle is conserved:

$$\epsilon = \frac{1}{2} m \{ (v_{\perp}^{(0)})^2 + (v_z^{(0)})^2 \} = \text{const.}$$

2. The projection of the angular momentum onto the z -axis is conserved, i.e.

$$L_z = m R_C^2 \dot{\varphi} = m R_C^2 \omega_C = \text{const.}$$

From formula (22.4) it follows that the magnetic moment produced by the particle moving in a circle is also conserved:

$$\mu = \frac{e}{2mc} L_z = \frac{e R_C^2 \omega_C}{2c} = \frac{1}{2} (m v_{\perp}^2) \frac{1}{H} = \frac{\epsilon_{\perp}}{H} = \text{const.}, \quad (39.10)$$

where $\epsilon_{\perp} = \frac{1}{2} m v_{\perp}^2$ is the kinetic energy of motion in the plane (xy). We shall make use of this important result in the next paragraph.

A constant uniform magnetic field possesses the property of focussing. Let a beam of particles emerge in different directions from a certain point with different initial velocities lying in the plane (xy). Since the cyclotron frequency does not depend on the initial velocity, the particles, having performed one revolution, will again come together at one point after the lapse of a time interval $T_C = 2\pi/\omega_C$.

If one now considers a beam of particles emerging in different directions, but having the same values of the initial velocity component $|v_{\perp}^{(0)}|$, then it can be seen that during a time T_C they will all traverse one turn of a helix. Its pitch, equal to $l = v_z^{(0)} T_C = v^{(0)} T_C \cos \alpha$, will in each case be different. Here α is the angle between the direction of the initial velocity and the z -axis. Hence the particles emerging from the initial point will not come together again at one point. If, however, the angle α is small, so that $\cos \alpha \approx 1$, then the pitch of the helix turns out to be the same for all the particles, and the beam is focussed.

Finally, let us consider the general case of the motion of a particle in electric and magnetic fields which are uniform and constant in time. We write the equation of motion in this case in the form

$$m \frac{d\mathbf{v}}{dt} = e\mathbf{E} + \frac{e}{c} \mathbf{v} \times \mathbf{H}.$$

We introduce a new unknown quantity \mathbf{V} , determined by the relation

$$\mathbf{V} = \mathbf{v} - \frac{c \mathbf{E} \times \mathbf{H}}{H^2}. \quad (39.11)$$

Substituting \mathbf{V} into the equation of motion, we find

$$\frac{d\mathbf{V}}{dt} = \frac{e}{m} \left(\mathbf{E} + \frac{1}{c} \mathbf{V} \times \mathbf{H} + \frac{(\mathbf{E} \times \mathbf{H}) \times \mathbf{H}}{H^2} \right).$$

Evaluating the vector product

$$(\mathbf{E} \times \mathbf{H}) \times \mathbf{H} = \mathbf{H}(\mathbf{H} \cdot \mathbf{E}) - E H^2,$$

we obtain

$$\frac{d\mathbf{V}}{dt} = \frac{e}{m} \frac{\mathbf{H}(\mathbf{H} \cdot \mathbf{E})}{H^2} + \frac{e}{mc} \mathbf{V} \times \mathbf{H}. \quad (39.12)$$

If the electric field is perpendicular to the magnetic field, so that $\mathbf{H} \cdot \mathbf{E} = 0$, then

$$\frac{d\mathbf{V}}{dt} = \frac{e}{mc} \mathbf{V} \times \mathbf{H}.$$

The above equation is the same as (39.3). Consequently, \mathbf{V} represents the velocity of motion of a particle in a circle in the plane perpendicular to the magnetic field \mathbf{H} with the cyclotron frequency. The components of the velocity \mathbf{V} are given by formulae analogous to (39.8), in which we assume that $\alpha = 0$.

In this case the total velocity of the particle is equal to

$$\mathbf{v} = \mathbf{V} + c \frac{\mathbf{E} \times \mathbf{H}}{H^2}$$

or, in the components,

$$v_x = V_x + cE_y/H = v_{\perp}^{(0)} \cos \omega_C t + cE_y/H, \quad (39.13)$$

$$v_y = V_y = -v_{\perp}^{(0)} \sin \omega_C t, \quad (39.14)$$

where $v_{\perp}^{(0)}$ is the initial value of the velocity in the plane perpendicular to the direction of the magnetic field \mathbf{H} . For the constant α equal to zero, the corresponding initial velocity of the particle is directed along the x -axis. From (39.13) it follows that v_x remains small in comparison with the velocity of light c , if the inequality $E_y \ll H$ holds.

The component of the velocity of the particle

$$v_D = c \frac{\mathbf{E} \times \mathbf{H}}{H^2} \quad (39.15)$$

is directed perpendicularly to both of the fields. Its absolute value is equal to

$$|v_D| = cE/H \quad (39.16)$$

and depends neither on the charge nor on the mass of the particle.

The motion of a particle in the direction of v_D is called "drift".

An obvious interpretation of the phenomenon of drift can be obtained from the following reasoning. Let a positively charged particle move on a circle in the plane (xy) perpendicular to the direction of the magnetic field \mathbf{H} ,

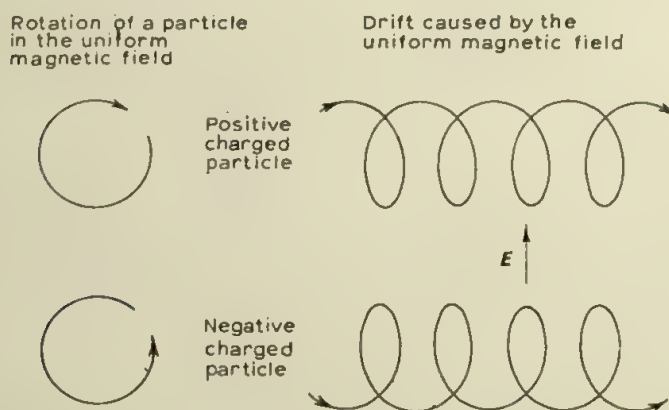


Fig. I.19

chosen to be the z -axis (in fig. I.19). The magnetic field is directed upward perpendicular to the plane of the drawing. Let, in addition, the electric field be directed along the y -axis. Then the electric field will accelerate the particle as it moves on the left semicircle, and decelerate it as it moves on the right semicircle. The circular trajectory will be distorted. The particle traverses the upper part of the circle with a larger velocity than the lower part. The magnetic field will bend the trajectory of the particle more in the lower part of the circle than in the upper one. Hence the projection of the path traversed by the particle onto the x -axis will be smaller in the lower part of the circle than in the upper one. As a result, after each revolution there arises a certain shift of the particle along the x -axis in its positive direction (in fig. I.19 from the left to the right), and the particle begins to move in the positive direction of the x -axis.

Similar reasoning for a negative particle leads to the same direction of drift.

Integrating the expressions (39.13) and (39.14) once more, we find the equations of the trajectory of the particle in parametric form:

$$x = \frac{v_{\perp}^{(0)}}{\omega_C} \sin \omega_C t + \frac{cEt}{H} + x_0, \quad (39.17)$$

$$y = \frac{v_{\perp}^{(0)}}{\omega_C} \cos \omega_C t + y_0. \quad (39.18)$$

The curve described by the particle is called a trochoid. The fixed parameters of the trochoid depend on the initial conditions. If it is assumed that at $t = 0$ the charged particle was at the origin, then $x_0 = 0$, $y_0 = -v_{\perp}^{(0)}/\omega_C$. In this case the form of the curve is determined only by the value of the initial velocity $v_{\perp}^{(0)}$. For $|v_{\perp}^{(0)}| > cE/H$ the upper curve of fig. I.20 is obtained, while for $|v_{\perp}^{(0)}| < cE/H$ the middle curve is obtained. The lower curve, a cycloid, corresponds to the case $v_{\perp}^{(0)} = -cE/H$.

If \mathbf{E} is not perpendicular to \mathbf{H} , then eq. (39.12) can be projected onto the plane perpendicular to \mathbf{H} and onto the z -axis. Then we find:

$$\frac{d\mathbf{V}_{\perp}}{dt} = \frac{e}{mc} \mathbf{V}_{\perp} \times \mathbf{H}, \quad \frac{dV_z}{dt} = \frac{e}{m} E_{\parallel},$$

where E_{\parallel} is the component of the electric field parallel to the magnetic field. In this case one has to write E instead of E_{\perp} in the drift velocity. The uniformly accelerated motion along the magnetic field under the action of the force eE_{\parallel} is superposed on the drift of the particle.

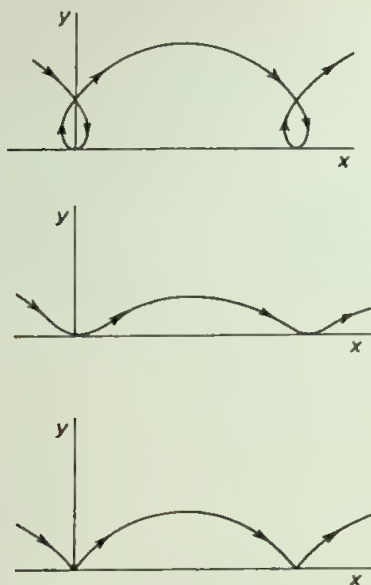


Fig. 1.20

§40. The motion of charged particles in slowly varying magnetic fields

Now we apply ourselves to the very important case of the motion of charged particles in magnetic fields varying in space and time. In the general case of varying fields the integration of the equations of motion turns out to be a very difficult problem. Hence we restrict ourselves to the case of fields varying slowly in time and space.

Consider the case where the magnetic field varies slowly in time, remaining uniform in space. Let a particle rotate with a cyclotron frequency ω_C in the plane perpendicular to the magnetic field. We assume that the change in the field per revolution is sufficiently small, i.e. that

$$T_C \left| \frac{\partial H}{\partial t} \right| \ll |H|, \quad (40.1)$$

where $T_C = 2\pi/\omega_C$.

For a time-varying field

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\frac{1}{c} \int \frac{\partial \mathbf{H}}{\partial t} \cdot d\mathbf{S}, \quad (40.2)$$

where the trajectory of the particle can be taken as the integration path. Multiplying (40.2) by the charge of the particle and assuming that during one revolution $\partial \mathbf{H} / \partial t$ remains constant, we can write:

$$e \oint \mathbf{E} \cdot d\mathbf{l} = -\frac{e}{c} \left| \frac{\partial \mathbf{H}}{\partial t} \right| S = -\frac{e}{c} \left| \frac{\partial \mathbf{H}}{\partial t} \right| \pi R_C^2. \quad (40.3)$$

The integral on the left of (40.3) represents the work done on the charge per revolution. It is equal to the increase of the kinetic energy of motion in the plane (xy), which we shall denote by $\Delta \epsilon_{\perp}$. Thus

$$\Delta \epsilon_{\perp} = -\frac{e}{c} \pi R_C^2 \left| \frac{\partial \mathbf{H}}{\partial t} \right| = \frac{|e| \pi R_C^2}{c} \left| \frac{\partial \mathbf{H}}{\partial t} \right| \quad (40.4)$$

The minus sign means that a particle with a negative charge moves in a direction opposite to the positive direction of the integration path.

Let us find the derivative $d\epsilon_{\perp}/dt$. On the basis of (40.4) and (39.10) we have

$$\begin{aligned} \frac{d\epsilon_{\perp}}{dt} &= \frac{\Delta \epsilon_{\perp}}{T_C} = \frac{|e| \pi R_C^2}{c T_C} \left| \frac{\partial \mathbf{H}}{\partial t} \right| = \frac{|e| \pi \omega_C R_C^2}{2\pi c} \left| \frac{\partial \mathbf{H}}{\partial t} \right| = \\ &= \mu \left| \frac{\partial \mathbf{H}}{\partial t} \right| = \mu \left| \frac{\partial H}{\partial t} \right| \end{aligned} \quad (40.5)$$

According to definition (39.10),

$$\frac{d\epsilon_{\perp}}{dt} = \mu \frac{\partial H}{\partial t} + H \frac{\partial \mu}{\partial t}. \quad (40.6)$$

Comparing eqs. (40.5) and (40.6) we see that, when the magnetic field varies slowly, the magnetic moment of the particle remains constant:

$$\frac{\partial \mu}{\partial t} = 0. \quad (40.7)$$

We can obtain the same result by considering the motion of a particle in a stationary magnetic field which varies slowly from point to point.

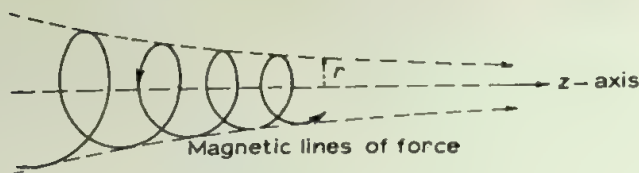


Fig. I.21

Let a magnetic field be symmetric about the z -axis and let its strength increase with increasing z . The field lines converge, as shown in fig. I.21. We shall assume that in a distance $\approx R_C$ the variation of the magnetic field is small, i.e.

$$R_C \left| \frac{\partial \mathbf{H}}{\partial z} \right| \ll |\mathbf{H}|. \quad (40.8)$$

Since the magnetic field varies along the z -axis, the radial component H_r of the field also differs from zero. From the continuity equation in cylindrical coordinates

$$\nabla \cdot \mathbf{H} = \frac{1}{r} \frac{\partial}{\partial r} (r H_r) + \frac{\partial H_z}{\partial z} = 0,$$

we have

$$\frac{\partial}{\partial r} (r H_r) = -r \frac{\partial H_z}{\partial z}.$$

Integrating, and disregarding the dependence of $\partial H_z / \partial z$ on the coordinate r on a circle of radius $r \sim R_C$, we find

$$H_r \approx -\frac{1}{2} r \frac{\partial H_z}{\partial z}.$$

Since, if the condition (40.8) is fulfilled, the component H_r is small in comparison with the component H_z for all $r < R_C$, it can be assumed that $|\mathbf{H}| \approx H_z$, i.e. that the field is directed at a small angle to the z -axis, and

$$H_r \approx -\frac{1}{2} r \frac{\partial |\mathbf{H}|}{\partial z}. \quad (40.9)$$

If the component of the magnetic field $H_z \neq 0$, then a particle moving in a circle in the (xy) plane with a cyclotron frequency is acted upon by a force in the direction of the z -axis. The particle will drift in this direction. Then the component v_z satisfies the equation of motion

$$m \frac{dv_z}{dt} = \frac{e}{c} v_{\perp} H_r = -\frac{e}{2c} v_{\perp} R_C \frac{\partial H}{\partial z} = -\frac{e}{2c} R_C^2 \omega_C \frac{\partial H}{\partial z} = -\mu \frac{\partial H}{\partial z}. \quad (40.10)$$

Whence in the usual way, multiplying (40.10) by v_z , we find

$$\frac{d}{dt} \left(\frac{1}{2} m v_z^2 \right) = -\mu \frac{\partial H}{\partial t}.$$

Since the total energy of the particle is conserved, we have

$$\frac{d}{dt} \left(\frac{1}{2} m v_z^2 + \frac{1}{2} m v_{\perp}^2 \right) = 0,$$

from which we obtain

$$\frac{d\epsilon_{\perp}}{dt} = \frac{d}{dt} \left(\frac{1}{2} m v_{\perp}^2 \right) = \mu \frac{\partial H}{\partial t}. \quad (40.11)$$

Comparing (40.11) with (40.6), we again arrive at the conclusion that the magnetic moment of the particle is conserved:

$$\frac{\partial \mu}{\partial t} = 0.$$

The conservation of the magnetic moment in a slowly varying non-uniform magnetic field leads to very important consequences. Since

$$\mu = \frac{\epsilon_{\perp}}{H}, \text{ and } R_C = \frac{cmv_{\perp}}{eH} = \frac{c}{eH} \sqrt{2m\epsilon_{\perp}} \sim \frac{1}{H},$$

the radii of the circles on which the particle moves decrease in the direction of increasing values of z (see fig. 1.21).

Let θ_0 be the angle formed by the velocity vector of the particle with the z -axis at a point z_0 , and let θ be the same angle at an arbitrary point. Then at the point $z = z_0$

$$\epsilon_{\perp} = \frac{1}{2} m v_{\perp}^2 = \frac{1}{2} m v_0^2 \sin^2 \theta_0 = \mu H(z_0),$$

where $H(z_0)$ is the value of the magnetic field at the point z_0 . At a certain point z the field strength is equal to $H(z)$, and

$$\epsilon_{\perp} = \mu H(z) = \frac{1}{2} m v_0^2 \sin^2 \theta.$$

Hence we have

$$\sin \theta = \sin \theta_0 \sqrt{\frac{H(z)}{H(z_0)}}.$$

As the particle moves along the z -axis and the field strength increases the angle increases. At a point z^* , where

$$\sqrt{\frac{H(z^*)}{H(z_0)}} \approx \frac{1}{\sin \theta_0},$$

$\sin \theta = 1$ and $v_{\perp} = v_0$. This means that the velocity component v_z of the particle reduces to zero. The particle cannot move farther than the point z^* , but is then reflected into the region $z < z^*$. The region $z \geq z^*$, which is impenetrable for particles with an initial velocity $v_{\perp} = v_0 \sin \theta_0$, is called a magnetic mirror.

The reflection of particles from the magnetic mirror plays a fundamental role in various electronic devices. E. Fermi put forward the idea of the acceleration of particles in cosmic rays as a result of reflection from magnetic mirrors. The role of the latter can be played by clouds of interstellar matter. If it is assumed that in the clouds of interstellar matter the magnetic field strength is larger than in the space separating them, then all particles confined between the clouds will be reflected from them as from magnetic mirrors. Assume that the clouds are moving towards each other with velocities v . Charged particles, colliding with the moving clouds and reflecting from them, change their velocity by $2v$ for each reflection. Calculations show that in cosmic conditions the velocities of the particles could reach enormous values.

In conclusion it should be stressed that all of the results obtained refer only to the motion of particles with velocities small in comparison with the velocity of light c . The motion with velocities comparable with the velocity of light c will be considered in Part II.

§41. The Lagrangian and Hamiltonian for a particle moving in an electromagnetic field

The equations of motion of a particle in an electromagnetic field

$$m \frac{d\mathbf{v}}{dt} = e \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right) \quad (41.1)$$

can be written in the form of Lagrange's equations, if the Lagrangian is introduced by the relation

$$L = \frac{1}{2} m v^2 - e\varphi + \frac{e}{c} \mathbf{A} \cdot \mathbf{v}. \quad (41.2)$$

We derive the Lagrange equations:

$$\frac{d}{dt} \frac{\partial L}{\partial \mathbf{v}} - \frac{\partial L}{\partial \mathbf{r}} = 0.$$

Clearly, for the generalized momentum we have

$$\mathbf{P} = \frac{\partial L}{\partial \mathbf{v}} = m \mathbf{v} + \frac{e}{c} \mathbf{A} = \mathbf{p} + \frac{e}{c} \mathbf{A}. \quad (41.3)$$

Correspondingly, the generalized force is

$$\begin{aligned} \left(\frac{\partial L}{\partial \mathbf{r}} \right)_{\mathbf{v}} &= -e \nabla \varphi + \frac{e}{c} \nabla (\mathbf{A} \cdot \mathbf{v}) = \\ &= -e \nabla \varphi + \frac{e}{c} \{ (\mathbf{v} \cdot \nabla) \mathbf{A} + \mathbf{v} \times (\nabla \times \mathbf{A}) \}. \end{aligned}$$

In calculating the partial derivative with respect to coordinates, \mathbf{v} was assumed to be constant.

Substituting these expressions into the Lagrange equations, we find

$$\frac{d}{dt} \left(\mathbf{p} + \frac{e}{c} \mathbf{A} \right) = -e \nabla \varphi + \frac{e}{c} (\mathbf{v} \cdot \nabla) \mathbf{A} + \frac{e}{c} \mathbf{v} \times (\nabla \times \mathbf{A}),$$

or

$$\frac{d\mathbf{p}}{dt} = -\frac{e}{c} \frac{d\mathbf{A}}{dt} + \frac{e}{c} (\mathbf{v} \cdot \nabla) \mathbf{A} + \frac{e}{c} \mathbf{v} \times (\nabla \times \mathbf{A}) - e \nabla \varphi.$$

However, according to (I.18) the total derivative with respect to time will be

$$\frac{d\mathbf{A}}{dt} = \frac{\partial \mathbf{A}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{A},$$

whence

$$\frac{d\mathbf{p}}{dt} = e \left\{ -\frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} - \nabla \varphi \right\} + \frac{e}{c} \mathbf{v} \times (\nabla \times \mathbf{A}),$$

or

$$\frac{d\mathbf{p}}{dt} = e \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right),$$

which is the same as (41.1).

We can also write the Hamiltonian for a particle:

$$\begin{aligned} H = \mathbf{P} \cdot \mathbf{v} - L &= \frac{\mathbf{P}}{m} \cdot \left(\mathbf{P} - \frac{e}{c} \mathbf{A} \right) - \frac{1}{2m} \left(\mathbf{P} - \frac{e}{c} \mathbf{A} \right)^2 + \\ &+ e\varphi - \frac{e\mathbf{A}}{mc} \cdot \left(\mathbf{P} - \frac{e}{c} \mathbf{A} \right) = \frac{1}{2m} \left(\mathbf{P} - \frac{e}{c} \mathbf{A} \right)^2 + e\varphi. \end{aligned} \quad (41.4)$$

In the case of a system of independent particles the Lagrangian and the Hamiltonian can be written in the form

$$\begin{aligned} L &= \sum \frac{1}{2} m_i v_i^2 - \sum e_i \varphi_i + \sum \frac{e_i}{c} \mathbf{A}_i \cdot \mathbf{v}_i = \\ &= \sum \frac{1}{2} m_i v_i^2 - \sum e_i \varphi_i + \frac{1}{c} \sum \mathbf{j}_i \cdot \mathbf{A}_i, \end{aligned} \quad (41.5)$$

and, correspondingly,

$$H = \sum \frac{1}{2m_i} \left(\mathbf{P}_i - \frac{e_i}{c} \mathbf{A}_i \right)^2 + \sum e_i \varphi_i, \quad (41.6)$$

where A_i and φ_i are the values of field potentials at the position of the i th particle. The summation is carried out over all particles of the system.

Later we shall need these expressions for the Lagrangian and the Hamiltonian.

It should be noted that, if one introduces the vector potential for a constant uniform magnetic field given by formula (19.16) into the expression for the generalized momentum P_i , then

$$P_i = p_i + \frac{e}{2c} \mathbf{H} \times \mathbf{r}, \quad (41.7)$$

and, consequently,

$$H = \sum \frac{1}{2m_i} \left(P_i - \frac{e_i}{2c} \mathbf{H} \times \mathbf{r} \right)^2 + \sum e_i \varphi_i \quad (41.8)$$

§42. The motion of a system of two charged particles and the radiation from them

Up to now we have considered the motion of particles in external fields. Now we shall discuss the problem of the motion of charged particles in the field produced by other particles.

Let us now consider the problem of the motion of two interacting charged particles*.

This problem can be solved by making use of the method of successive approximations. Namely, assuming the energy losses of the particles due to radiation to be small, the trajectories of the particles can be calculated in the first approximation. Knowing them, one can then find the radiation of the system.

Let the charged particles have masses m_1 and m_2 , and charges e_1 and e_2 . The potential energy of the system can be written in the form

$$U = e_1 \varphi(r),$$

* In what follows we shall confine ourselves to a brief exposition of the problem, since it appears to be a particular case of the two-body problem considered in detail in courses in classical mechanics. A detailed exposition of this and other problems touched upon in this paragraph can be found in: L.D. Landau and E.M. Lifshitz, *Mechanics* (Pergamon, Oxford, 1960), and H. Goldstein, *Classical mechanics* (Addison-Wesley, Reading, Mass., 1950).

where $\varphi(r)$ is the potential of the field produced by the charge e_2 located at a distance r from the charge e_1 .

The Lagrangian of the system can be written

$$L = \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 - e_1 \varphi(r), \quad (42.1)$$

where $\mathbf{v}_1 = \dot{\mathbf{r}}_1$, $\mathbf{v}_2 = \dot{\mathbf{r}}_2$, while \mathbf{r}_1 and \mathbf{r}_2 are radius-vectors drawn to the particles from an arbitrary origin.

Since the potential energy depends only on the distance between the charges, i.e. $U = U(|\mathbf{r}|)$, where

$$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2, \quad (42.2)$$

it is convenient to pass over to the centre-of-mass system.

That is to say, we place the origin at the centre of mass, i.e. the spatial point whose radius-vector with respect to an arbitrary system of coordinates is expressed by the formula

$$\mathbf{R} = \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{m_1 + m_2}. \quad (42.3)$$

From (42.2) and (42.3) we find

$$\begin{aligned} \mathbf{r}_1 &= \frac{m_2}{m_1 + m_2} \mathbf{r} + \mathbf{R}, \\ \mathbf{r}_2 &= -\frac{m_1}{m_1 + m_2} \mathbf{r} + \mathbf{R}, \end{aligned}$$

and, correspondingly, the velocities of the particles are

$$\begin{aligned} \mathbf{v}_1 &= \frac{m_2}{m_1 + m_2} \dot{\mathbf{r}} + \dot{\mathbf{R}} = \frac{m_2}{m_1 + m_2} \mathbf{v}_0 + \dot{\mathbf{R}}, \\ \mathbf{v}_2 &= -\frac{m_1}{m_1 + m_2} \dot{\mathbf{r}} + \dot{\mathbf{R}} = -\frac{m_1}{m_1 + m_2} \mathbf{v}_0 + \dot{\mathbf{R}}, \end{aligned} \quad (42.4)$$

where $\mathbf{v}_0 = \dot{\mathbf{r}}$ is the relative velocity of the particles, and $\dot{\mathbf{R}}$ is the velocity of the centre of mass.

Inserting these values of the velocities into the Lagrangian, we find

$$L = \frac{1}{2} M (\dot{\mathbf{R}})^2 + \frac{1}{2} \mu (\dot{\mathbf{r}})^2 - U(|\mathbf{r}|), \quad (42.5)$$

where $M = m_1 + m_2$ is the mass of the system, while the quantity

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (42.6)$$

is called the reduced mass.

The coordinates of the centre of mass \mathbf{R} are cyclic. Hence the corresponding generalized momentum is conserved:

$$\frac{\partial L}{\partial \mathbf{R}} = (m_1 + m_2) \mathbf{R} = \text{const}. \quad (42.7)$$

The centre of mass moves with a constant velocity (in particular, it can remain at rest).

In order to investigate the relative motion of the charges we introduce coordinates whose origin coincides with the centre of mass. Since the potential energy of the interaction depends only on the distance r , the field has spherical symmetry. Therefore the coordinate corresponding to arbitrary rotation of the system is cyclic. This means that the angular momentum conservation law holds:

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} = \text{const}. \quad (42.8)$$

Taking the scalar product of this expression and the radius-vector \mathbf{r} , we have

$$\mathbf{L} \cdot \mathbf{r} = 0,$$

so that the motion takes place in the plane perpendicular to the vector \mathbf{L} .

Choosing the direction of the vector \mathbf{L} to be the z -axis, and the plane in which the motion takes place to be the plane $z = 0$, we can introduce the polar coordinates r, ψ and rewrite the Lagrangian referring to the relative motion in the form

$$L_{\text{rel}} = \frac{1}{2} \mu (\dot{r}^2 + r^2 \dot{\psi}^2) - e_1 \varphi(r). \quad (42.9)$$

The Lagrangian L_{rel} is formally the same as the Lagrangian of one particle having mass μ and moving in the external force field $e_1 \varphi$. The coordinate ψ is cyclic, and the generalized momentum

$$p_\psi = \mu r^2 \dot{\psi} = \text{const} \quad (42.10)$$

corresponds to it. Calculating L_z in polar coordinates, we have, obviously,

$$L_z = [\mathbf{r} \times \mathbf{p}]_z = \mu(xy - yx) = \mu r^2 \dot{\psi}, \quad (42.11)$$

so that the equality

$$p_\psi = L_z = \text{const} = L$$

expresses the angular momentum conservation law.

Since the Lagrangian does not depend on time, the energy conservation law holds:

$$E = \frac{1}{2} \mu (\dot{r}^2 + r^2 \dot{\psi}^2) + e_1 \varphi(r) = \text{const}. \quad (42.12)$$

Expressing $\dot{\psi}$ from (42.11) in terms of L_z , we find

$$E = \frac{1}{2} \mu \dot{r}^2 + \frac{L^2}{2\mu r^2} + e_1 \varphi = \frac{1}{2} \mu \dot{r}^2 + V(r), \quad (42.12')$$

where the quantity $V(r)$, called the effective potential energy, is equal to

$$V(r) = e_1 \varphi(r) + \frac{L^2}{2\mu r^2}. \quad (42.13)$$

The energy (42.12') is formally the same as the energy of a particle moving uniformly in a field with potential energy $V(r)$.

The character of the relative motion of the charged particles is defined by the form of the function $\varphi(r)$.

Here we shall consider the case of oppositely charged particles. Then

$$\varphi = -e_2/r, \quad (42.14)$$

where e_2 is the charge of the second particle.

The variation of the energy as a function of r in this case is shown in fig. I.22. The dashed lines show the curves $-e_1 e_2/r$ and $L^2/2\mu r^2$. The straight line $E = 0$ is given by a dotted line. The shape of the curve $V(r)$ and the position of the minimum depend on the value of the angular momentum L .

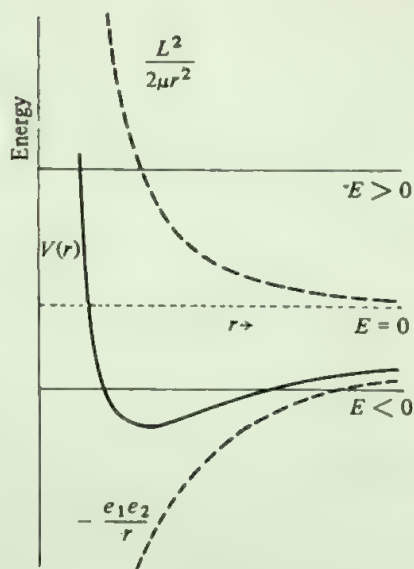


Fig. 1.22

Rewriting the formula (42.12') in the form

$$\dot{r}^2 = \frac{2}{\mu}(E - V(r)), \quad (42.15)$$

we see that the allowed region of the relative motion of the particles depends on the relation between E and $V(r)$. The regions in which $V(r) > E$ are forbidden, whereas those in which $E > V(r)$ are allowed. If the roots of the equation

$$E = V(r) \quad (42.16)$$

exist, they determine the radii r_0 at which the radial velocity reduces to zero. If $E \geq 0$, then the allowed region of motion extends from the infinitely distant region $r \rightarrow \infty$ up to the radius corresponding to the minimum distance between the particles and determined by the crossing of the straight line E with the curve $V(r)$.

If $E < 0$, then there are two radii corresponding to the smallest and largest distance between the charges. The motion is performed between these points.

In order to find the trajectory, the time should be eliminated from (42.12') and (42.11).

We then have

$$d\psi = \frac{L}{\mu r^2} dt = \frac{L}{\mu r^2} \frac{dr}{\sqrt{\frac{2}{\mu}(E - e_1\varphi(r)) - \frac{L^2}{\mu^2 r^2}}}.$$

By integration we find

$$\psi = \int \frac{dr}{r^2 \sqrt{\frac{2\mu}{L^2}(E - e_1\varphi) - \frac{1}{r^2}}} - \psi_0. \quad (42.17)$$

In particular, for the motion in the attractive Coulomb field the substitution for $\varphi(r)$ from (42.14) gives

$$\begin{aligned} \psi &= \int \frac{dr}{r^2 \sqrt{\frac{2\mu}{L^2} \left(E + \frac{e_1 e_2}{r} \right) - \frac{1}{r^2}}} = - \int \frac{du}{\sqrt{\frac{2\mu E}{L^2} + \frac{2e_1 e_2 \mu}{L^2} u - u^2}} = \\ &= - \arccos \frac{\frac{L^2 u}{e_1 e_2 \mu} - 1}{\sqrt{1 + \frac{2EL^2}{\mu(e_1 e_2)^2}}} - \psi_0, \end{aligned}$$

where

$$u = 1/r.$$

We find for the equation of the trajectory:

$$\left(\frac{L^2}{\mu e_1 e_2} \right) \frac{1}{r} = \left[1 + \sqrt{1 + \frac{2EL^2}{\mu(e_1 e_2)^2}} \cos(\psi + \psi_0) \right]. \quad (42.18)$$

Comparing this formula with the general equation of conic sections

$$p/r = 1 + \epsilon \cos(\psi + \psi_0),$$

where ϵ is the eccentricity, we see that the motion of the charges takes place over a conic section with eccentricity

$$\epsilon = \sqrt{1 + \frac{2EL^2}{\mu(e_1e_2)^2}}. \quad (42.19)$$

The parameter of the conic section is $p = L^2/\mu e_1 e_2$. Depending on the sign of E we have

$$E > 0, \quad \epsilon > 1 \quad \text{hyperbola}$$

$$E = 0, \quad \epsilon = 1 \quad \text{parabola}$$

$$E < 0, \quad \epsilon < 1 \quad \text{ellipse}$$

$$E = -\frac{\mu(e_1e_2)^2}{2L^2}, \quad \epsilon = 0 \quad \text{circle}.$$

The solution (42.18) determines the motion of each of the particles. Namely, the particles move on trajectories representing conic sections with the foci located at the centre of mass of the system.

This solution does not differ from the corresponding solution of the Kepler problem on the motion of planets. The values of the quantities characteristic of conic sections are given in the books on mechanics cited.

Let us now find the radiation from a system of two charges. Here we confine ourselves to the case of periodic motion, i.e. to the case of attraction for $E < 0$. The energy emitted during a period T is, according to (27.9) and (28.11), equal to

$$-(\Delta E)_T = \frac{2}{3c^3} \int (\ddot{\mathbf{d}})^2 dt = \frac{2}{3c^3} \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right)^2 \int_0^T F^2 dt,$$

where

$$\mathbf{F} = \frac{e_1 e_2}{r^3} \mathbf{r}.$$

Instead of integrating over a period, we can integrate over the angle ψ by means of (42.11). Expressing r in terms of ψ according to formula (42.11), we obtain

$$\begin{aligned}
-(\Delta E)_T &= \frac{2}{3c^3} \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right)^2 (e_1 e_2)^2 \int \frac{dt}{r^4} = \\
&= \frac{2}{3c^3} \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right)^2 (e_1 e_2)^2 \frac{\mu}{L} \int_0^{2\pi} \frac{d\psi}{r^2} = \\
&= \frac{2}{3c^3} \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right)^2 \frac{(e_1 e_2 \mu)^4}{\mu L^5} \int_0^{2\pi} [1 + \epsilon \cos(\psi + \psi_0)]^2 d\psi = \\
&= \frac{2}{3c^3} \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right)^2 \frac{(e_1 e_2 \mu)^4 \pi}{\mu L^5} (2 + \epsilon^2). \quad (42.20)
\end{aligned}$$

Thus, a charge moving in a closed orbit continuously emits energy. In particular, for motion in a circular orbit $\epsilon = 0$, $L^2 = -\mu(e_1 e_2)^2/2E$, and formula (42.20) reduces to the simpler form:

$$-(\Delta E)_T = \frac{16\pi\sqrt{2\mu}}{3c^3 e_1 e_2} \left(\frac{e_1}{m_1} - \frac{e_2}{m_2} \right)^2 |E|^{5/2} \quad (42.21)$$

The energy loss due to radiation leads to the transformation of the circular orbit into a spiral.

Numerical calculation shows that a system consisting of an electron rotating about a proton emits its energy during a time $T \approx 10$ sec for linear dimensions of the orbit $\sim 10^{-8}$ cm.

We see that a planetary atomic model contradicts the laws of classical electrodynamics. Later, in expounding quantum mechanics, the cause of this contradiction will be explained. It turns out that the instability of the planetary atomic model is an illustration of the general statement on the inapplicability of the laws of classical electrodynamics to the consideration of intra-atomic phenomena.

§43. The scattering of particles and associated bremsstrahlung

We shall now consider processes in which the particles are in relative motion with either a repulsive or attractive interaction, but having an energy $E > 0$. From general considerations associated with the shape of the curve $V(r)$ (see fig. 1.22) it is clear that in both cases the motion will take place in an open orbit.

For simplicity we assume that one of the charged particles is at rest with respect to the laboratory system of coordinates. We call it the target. The other, incident particle, which is scattered, moves relative to the first one. At a sufficiently large distance, when the interaction between the particles can be disregarded, the motion of the incident particle is rectilinear. We assume that the velocity v_0 of this motion is given. When the particles approach each other the incident particle is deflected from its rectilinear motion, and the target particle, which was initially at rest, acquires momentum and is set in motion.

The particles are said to have undergone collision with each other, as a result of which they are scattered.

For reasons which will be clear from what follows, the investigation of the process of scattering leads to most important information about the nature of the interaction between the particles.

The study of scattering processes appears at present to be the basic experimental method of nuclear physics. The investigation of the interaction of particles, for example, fast electrons or protons with nuclei, is usually carried out in the following way. A beam of particles having well defined properties and a known velocity is incident upon a sample of matter containing particles of another kind. From observations on the beam of scattered particles one can draw conclusions about the nature of the interaction which led to the scattering.

In such an experimental set-up, scattering has the character of a bulk process. The behaviour of a beam, usually containing an enormous number of particles, is observed. However, the process is based on the individual interaction between the particle to be scattered and the target particle. Hence the process of scattering must be characterized by a quantity which depends neither on the properties of the incident beam nor on the properties of the target material, for example, its density, but which is determined solely by the interaction between a particle which is to be scattered and a target particle.

We characterize the incident particle beam by its intensity or the density of the flux of particles $I_0 = nv_0$, where n is the number of particles per unit volume of the beam, and v_0 is their velocity. I_0 is obviously equal to the number of particles traversing 1 cm^2 of the cross section of the beam per sec.

We choose the position of the target particle to be the origin. Let dN scattered particles per unit time enter the solid angle $d\Omega$ about the scattering centre. We define the basic quantity characterizing the process of scattering — the differential cross section for the scattering, $d\sigma$ — as the ratio

$$d\sigma = dN/I_0. \quad (43.1)$$

The cross section obviously has the dimensions of area. The number of particles scattered from the beam by the target volume V into the solid angle $d\Omega$ during the time dt is equal to

$$dN_{\text{tot}} = d\sigma I_0 \rho V dt, \quad (43.2)$$

where ρ is the number density of the target particles (ρV is the total number of target particles).

If the scattering occurs without a change in the energy of the scattered particles, then, multiplying the numerator and the denominator of (43.1) by the energy ϵ of the particle, we can write the differential cross section in the form

$$d\sigma = dI/I_{0\epsilon}, \quad (43.3)$$

where dI is the flux of energy carried by the particles per unit time in the solid angle $d\Omega$, and $I_{0\epsilon}$ is the intensity of the energy flux in the incident beam. Formula (43.3) is the same as the definition of cross section given in §36.

In addition to the differential cross section, scattering is also characterized by the total cross section

$$\sigma = \int d\sigma, \quad (43.4)$$

where the integration is carried out over all possible values of solid angle.

The quantity $d\sigma$, expressed in terms of directly measurable quantities, can be connected with parameters characterizing the individual collision process.

Consider an individual act of collision between two particles. We confine ourselves to the case where the internal energy of the particles remains unchanged. Such collisions are called elastic. It should not be thought that in elastic collisions the energy of the scattered particle does not change. The target particle receives from the scattered one a certain momentum and energy, whose value depends on the mass ratio of the particles.

In inelastic collisions, for example collision of an electron with an ion, there occurs an additional transfer of energy to the ion, whose internal state changes. Such processes are more complicated and will be considered in Part V.

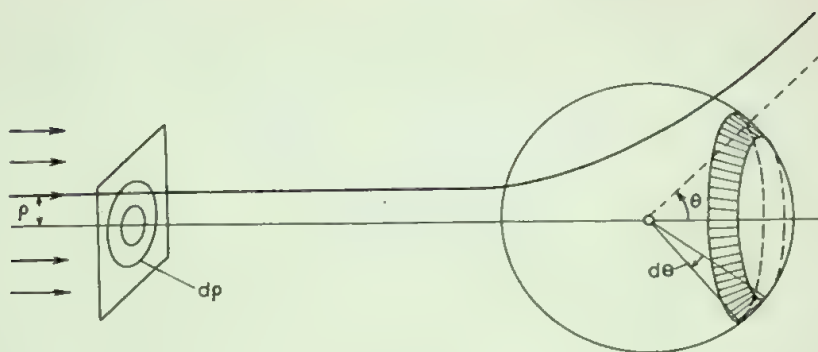


Fig. 1.23

At first we shall refer the process of scattering of two particles to the centre-of-mass system. According to the results of the preceding paragraph, the problem of the relative motion of particles in the centre-of-mass system reduces to the problem of the motion of one particle with reduced mass μ relative to the motionless force centre located at the centre of mass.

At a large distance from the centre of mass let the incident particle move rectilinearly with a velocity v_0 . Its energy and angular momentum are respectively equal to

$$E = \frac{\mu v_0^2}{2}, \quad L = \mu v_0 \rho = \rho \sqrt{2\mu E}, \quad (43.5)$$

where ρ is the distance between the force centre and the straight line along which the particle would pass by if there were no interaction. The quantity ρ is called the impact parameter.

For given values of E and ρ the trajectory is completely determined. We characterize the process of scattering by the scattering (deflection) angle θ , representing the angle between the directions of motion of the particle at a large distance from the centre before and after the scattering (see fig. 1.23). It is obvious that the angle θ is the complement of the angle ψ_0 between the asymptotes of the trajectory (fig. 1.24). Because of the symmetry of the force field and the distribution of the trajectories relative to the axis of the beam, the number of scattered particles and the cross section for scattering depend only on the angle θ and not on the azimuthal angle.

The solid angle $d\Omega$ can therefore be written in the form

$$d\Omega = 2\pi \sin \theta \, d\theta.$$

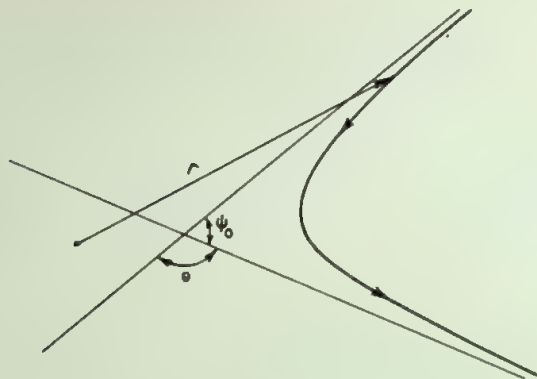


Fig. 1.24

The number of particles scattered into the solid angle $d\Omega$ is correspondingly equal to

$$dN = 2\pi I_0 \sigma(\theta) \sin \theta \, d\theta = I_0 \, d\sigma, \quad (43.6)$$

where

$$d\sigma = 2\pi \sigma(\theta) \sin \theta \, d\theta.$$

Since the trajectory of the scattered particle is uniquely determined by the energy and the impact parameter (E and ρ), there corresponds to each scattering angle a trajectory with a definite value of the impact parameter. Hence it follows that the number of particles scattered at a given angle θ is equal to the number of particles having given value of the impact parameter at an infinitely large distance from the centre.

In other words, all particles having a value of the impact parameter lying in the interval $\rho, \rho + d\rho$ are scattered into the solid angle $d\Omega$. Hence the number of scattered particles can also be written in the form

$$-dN = I_0 2\pi \rho \, d\rho, \quad (43.7)$$

where $2\pi \rho \, d\rho$ is the area of the ring shown in fig. 1.23 on the left.

Comparing (43.6) and (43.7) we find

$$\sigma(\theta) = -\frac{\rho}{\sin \theta} \frac{d\rho}{d\theta} = \frac{\rho}{\sin \theta} \left| \frac{d\rho}{d\theta} \right| \quad (43.8)$$

Since to large values of ρ there correspond small deflections θ , and since the cross section, by definition, must have a positive value, we have written the absolute value of the derivative. Integrating formula (43.8) on the left with respect to scattering angles in the interval from θ to π , and on the right with respect to the corresponding values of ρ , i.e. in the interval from $\rho(\theta)$ to zero, we obtain the important relation

$$\frac{1}{2}\rho^2(\theta) = \int_{\theta}^{\pi} \sigma(\theta) \sin \theta \, d\theta. \quad (43.9)$$

In order to calculate the differential cross section it is necessary to establish a relation between the impact parameter ρ and the scattering angle θ . For this it is sufficient to calculate the trajectory and find the dependence of the angle ψ_0 between the asymptotes on the parameter ρ .

General formula (42.17) gives

$$\begin{aligned} \psi_0 &= \int \frac{L}{r^2} \frac{dr}{\sqrt{2\mu(E - U(r)) - \frac{L^2}{r^2}}} = \int \frac{dr}{r^2 \sqrt{\frac{2\mu(E - U(r))}{L^2} - \frac{1}{r^2}}} = \\ &= \int \frac{dr}{r^2 \sqrt{\frac{1}{\rho^2} \left(1 - \frac{U(r)}{E}\right) - \frac{1}{r^2}}} \end{aligned} \quad (43.10)$$

The limits of the integral are determined from the following considerations. The angle ψ_0 represents the change in the angle ψ when the particle describes all its trajectory. The trajectory has two branches: the motion of the particle from infinity to the point of closest approach r_0 , and its motion away from the point r_0 to infinity (see fig. I.24). The trajectory of the particle is always symmetric relative to the point of closest approach r_0 . This follows from the reversibility of the process of scattering: the particle moving in the opposite direction must move on the same asymptote on which the particle moves forward (see fig. I.24). Hence the integral over the trajectory can be written in the form of the sum of two equal integrals taken in the range from r_0 to infinity:

$$\psi_0 = 2 \int_{r_0}^{\infty} \frac{dr}{r^2 \sqrt{\frac{1}{\rho^2} \left(1 - \frac{U(r)}{E}\right) - \frac{1}{r^2}}}, \quad (43.11)$$

where r_0 is the root of the equation

$$E = V(r_0) = U(r_0) + \frac{\rho^2 E}{r_0^2}. \quad (43.12)$$

Since the scattering angle is $\theta = \pi - \psi_0$, formulae (43.11) and (43.12) connect the value of θ sought with the parameters of the collision — i.e. the quantities E and ρ — for an arbitrary form of the potential U depending only on the distance r .

Let us consider, in particular, the case of Coulomb repulsion:

$$U = |e_1 e_2| / r.$$

Expressions (43.12) and (43.11) then take the forms

$$E = \frac{|e_1 e_2|}{r_0} + \frac{\rho^2 E}{r_0^2}, \quad (43.13)$$

or

$$\frac{1}{r_0} = -\frac{|e_1 e_2|}{2E\rho^2} + \sqrt{\left(\frac{|e_1 e_2|}{2E\rho^2}\right)^2 + \frac{1}{\rho^2}} \quad (43.14)$$

and

$$\psi_0 = 2 \int_{r_0}^{\infty} \frac{dr}{r^2 \sqrt{\frac{1}{\rho^2} \left(1 - \frac{|e_1 e_2|}{Er}\right) - \frac{1}{r^2}}} \quad (43.15)$$

Introducing the new variable $z = 1/r$, we obtain

$$\begin{aligned} \psi_0 &= 2 \int_0^{z_0} \frac{dz}{\sqrt{\frac{1}{\rho^2} - \frac{|e_1 e_2|}{\rho^2 E} z - z^2}} = -2 \arccos \frac{\frac{|e_1 e_2|}{2\rho^2 E} + z}{\sqrt{\left(\frac{|e_1 e_2|}{2\rho^2 E}\right)^2 + \frac{1}{\rho^2}}} \bigg|_0^{z_0} = \\ &= 2 \arccos \frac{\frac{|e_1 e_2|}{2\rho^2 E}}{\sqrt{\left(\frac{|e_1 e_2|}{2\rho^2 E}\right)^2 + \frac{1}{\rho^2}}}. \end{aligned} \quad (43.16)$$

In the calculation we substituted the value $z_0 = 1/r_0$ from expression (43.14). Further we find

$$\cos \frac{1}{2}\psi_0 = \frac{\frac{|e_1 e_2|}{2\rho^2 E}}{\sqrt{\left(\frac{|e_1 e_2|}{2\rho^2 E}\right)^2 + \frac{1}{\rho^2}}},$$

or

$$\rho = \frac{|e_1 e_2|}{2E} \tan \frac{1}{2}\psi_0 = \frac{|e_1 e_2|}{2E} \tan \frac{1}{2}(\pi - \theta) = \frac{|e_1 e_2|}{2E} \cot \frac{1}{2}\theta. \quad (43.17)$$

Formula (43.17) gives the relation sought between ρ and θ . Substituting the value of ρ into (43.8), we have

$$\sigma(\theta) = \frac{\rho}{\sin \theta} \left| \frac{d\rho}{d\theta} \right| = \frac{1}{16} \frac{(e_1 e_2)^2}{E^2} \frac{1}{\sin^4 \frac{1}{2}\theta}. \quad (43.18)$$

Formula (43.18) gives the differential cross section for the scattering of particles which repel each other according to Coulomb's law. Since $\sigma \sim (e_1 e_2)^2$, it is clear that an identical result is obtained in the case of attraction (for $E > 0$). The expression found for the cross section is called the Rutherford formula. It was obtained in connection with experiments on the scattering of α -particles by atoms which allowed Rutherford to establish the nuclear structure of atoms.

Formula (43.11) can be inverted and the form of the function $U(r)$ can be found from the results of experiment. For this it is necessary to consider (43.11) as an integral equation with respect to the unknown function $U(r)$. Under certain (very general) assumptions about the form of the function $U(r)$ eq. (43.11) can be solved*.

We write θ in the form

$$\theta = \pi - \psi_0 = \pi - \int_{r_0}^{\infty} \frac{2\rho dr}{r \sqrt{\left(1 - \frac{U(r)}{E}\right) r^2 - \rho^2}}. \quad (43.19)$$

As for $U(r)$ we shall assume that it decreases with increasing r for $r \rightarrow \infty$. We introduce a new function

* We follow the work of O.B. Firsov, Zhur. Exp. i Teoret. Fiz. 24 (1953) 279.

$$F(r) = \left(1 - \frac{U(r)}{E}\right) r^2, \quad (43.20)$$

and go over from integration with respect to r to integration with respect to F . Since

$$2 \frac{dr}{r} = 2 d \ln r = \frac{d \ln r^2}{dF} dF,$$

and, using (43.12),

$$F(r_0) = \left(1 - \frac{U(r_0)}{E}\right) r_0^2 = \rho^2, \quad F(\infty) = \infty \quad (43.21)$$

we have

$$\theta = \pi - \int_{F(r_0)}^{F(\infty)} \rho \frac{d \ln r^2}{dF} \frac{dF}{\sqrt{F - \rho^2}}. \quad (43.22)$$

Writing

$$\frac{d \ln r^2}{dF} \equiv \frac{d}{dF} \ln \left(\frac{r^2}{F} \right) + \frac{d \ln F}{dF},$$

we find, substituting for the bounds $F(r_0)$ and $F(\infty)$ the values (43.21),

$$\theta = \pi + \int_{\rho^2}^{\infty} \rho \frac{d}{dF} \ln \left(\frac{F}{r^2} \right) \frac{dF}{\sqrt{F - \rho^2}} - \int_{\rho^2}^{\infty} \rho \frac{d \ln F}{dF} \frac{dF}{\sqrt{F - \rho^2}}.$$

The last integral is easy to calculate, because

$$\int_{\rho^2}^{\infty} \frac{dF}{F \sqrt{F - \rho^2}} = \frac{2}{\rho} \arctan \frac{\sqrt{F - \rho^2}}{\rho} \bigg|_{\rho^2}^{\infty} = \frac{\pi}{\rho}.$$

Hence

$$\theta = \int_{\rho^2}^{\infty} \rho \frac{d \ln \left(\frac{F'}{r'^2} \right)}{dF'} \frac{dF'}{\sqrt{F' - \rho^2}}. \quad (43.23)$$

We multiply both sides of (43.23) by the factor $1/\sqrt{\rho^2 - F}$ and integrate with respect to the parameter ρ in the range from $\rho = \sqrt{F}$ to $\rho \rightarrow \infty$. Then we have

$$\int_{\sqrt{F}}^{\infty} \frac{\theta(\rho) d\rho}{\sqrt{\rho^2 - F}} = \int_{\sqrt{F}}^{\infty} \rho d\rho \int_{\rho^2}^{\infty} \frac{d \ln \left(\frac{F'}{r'^2} \right)}{dF'} \frac{dF'}{\sqrt{(\rho^2 - F)(F' - \rho^2)}}.$$

We change the order of integration with respect to ρ and F' , writing

$$\begin{aligned} \int_{\sqrt{F}}^{\infty} \rho d\rho \int_{\rho^2}^{\infty} \frac{d \ln \left(\frac{F'}{r'^2} \right)}{dF'} dF' &= \\ &= \int_F^{\infty} \frac{d \ln \left(\frac{F'}{r'^2} \right)}{dF'} dF' \int_{\sqrt{F'}}^{\sqrt{F}} \frac{\rho d\rho}{\sqrt{-\rho^4 + \rho^2(F + F') - FF'}}. \end{aligned}$$

Further,

$$\begin{aligned} \frac{1}{2} \int_{\rho=\sqrt{F'}}^{\rho=\sqrt{F}} \frac{d\rho^2}{\sqrt{-\rho^4 + \rho^2(F + F') - FF'}} &= \\ &= -\frac{1}{2} \arcsin \frac{(F + F') - \rho^2}{\sqrt{(F + F')^2 - 4FF'}} \Big|_{\rho=\sqrt{F'}}^{\rho=\sqrt{F}} = \frac{1}{2} \pi. \end{aligned}$$

The fact that, as a result of integration with respect to ρ , a factor independent of F' is obtained allows us to reduce the integration with respect to F' to an elementary transformation. In order to obtain a more explicit result we return, before making the substitution of the limits, to the variable r , writing

$$\begin{aligned} \int_{F(r)}^{F(\infty)} \frac{d}{dF'} \ln \left(\frac{F'}{r'^2} \right) dF' &= \int_r^{\infty} \frac{d}{dr'} \ln \left(\frac{F'}{(r')^2} \right) dr' = \\ &= \ln \left(1 - \frac{U(r')}{E} \right) \Big|_r^{\infty} = -\ln \left(1 - \frac{U(r)}{E} \right) = -\ln \frac{F}{r^2}. \end{aligned}$$

According to the above assumption we put $U \rightarrow 0$ as $r \rightarrow \infty$.

As a result we arrive at the equality

$$\pi \ln \frac{\sqrt{F}}{r} = - \int_{\sqrt{F}}^{\infty} \frac{\theta(\rho) d\rho}{\sqrt{F} \sqrt{\rho^2 - F}}, \quad (43.24)$$

or

$$\frac{\sqrt{F}}{r} = \exp \left\{ - \frac{1}{\pi} \int_{\sqrt{F}}^{\infty} \frac{\theta(\rho) d\rho}{\sqrt{\rho^2 - F}} \right\}. \quad (43.25)$$

The last expression contains integration only with respect to ρ . If the dependence $\sigma(\theta)$ is known, then by virtue of (43.19) the functional dependence $\theta(\rho)$ is also given. Hence formula (43.25) gives the function $F(r)$ in implicit form.

A particularly simple result is obtained for the scattering of particles with a large energy, for which $U(r) \ll E$ for all values of r . Then formula (43.23) can be rewritten, making use of the equalities

$$\ln \frac{\sqrt{F}}{r} = \ln \frac{\sqrt{\left(1 - \frac{U(r)}{E}\right) r^2}}{r} \approx - \frac{U(r)}{2E},$$

$$\int_{\sqrt{F}}^{\infty} \frac{\theta(\rho) d\rho}{\sqrt{F} \sqrt{\rho^2 - F}} = \int_{\sqrt{\left(1 - \frac{U(r)}{E}\right) r^2}}^{\infty} \frac{\theta(\rho) d\rho}{\sqrt{\rho^2 - \left(1 - \frac{U(r)}{E}\right) r^2}} \approx \int_r^{\infty} \frac{\theta(\rho) d\rho}{\sqrt{\rho^2 - r^2}}.$$

and we find the potential energy sought:

$$U(r) \approx \frac{2E}{\pi} \int_r^{\infty} \frac{\theta(\rho) d\rho}{\sqrt{\rho^2 - r^2}}. \quad (43.26)$$

The above calculations of the scattering cross section have been carried out in the centre-of-mass system. For their practical use it is necessary to make a transition to the laboratory system.

The importance of this transition is seen from the general picture of the process of scattering from the point of view of the two coordinate systems. In

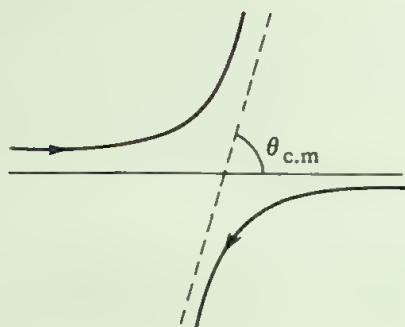


Fig. I.25

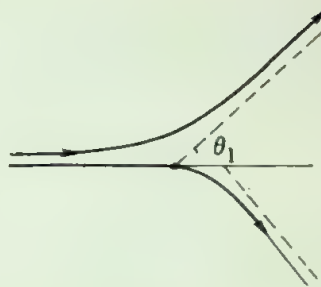


Fig. I.26

the centre-of-mass system each of the particles, the incident and the target one, move on a trajectory determined by formula (43.10). From the fact that the total momentum of the system of two particles, referred to the centre-of-mass system, is equal to zero it follows that the two particles move in opposite directions with equal momenta (fig. I.25). On the contrary, from the standpoint of the laboratory system the incident and target particles are not equivalent. Before collision the target particle was at rest. After collision it is set in motion (fig. I.26).

Our problem is to establish a relation between the angles $\theta_{c.m.}$ in the centre-of-mass system and θ_1 in the laboratory system. For this we put in the same diagram the velocities of the scattered particle after collision in the centre-of-mass ($\mathbf{v}'_{c.m.}$) and laboratory (\mathbf{v}'_1) systems. It is obvious that equalities (42.4) and (42.7) hold. The vector \mathbf{R} coincides with the direction of motion of the incident particle. Hence the angles sought for are respectively equal to (fig. I.27):

$$\begin{aligned}\theta_{c.m.} &= \angle(\mathbf{v}'_{c.m.}, \dot{\mathbf{R}}), \\ \theta_1 &= \angle(\mathbf{v}'_1, \dot{\mathbf{R}}),\end{aligned}\tag{43.27}$$

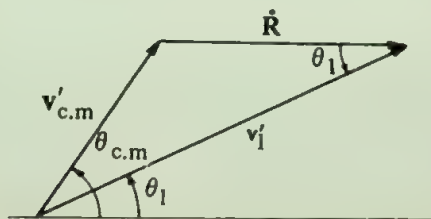


Fig. I.27

where the prime denotes that the values of the velocity are taken after collision. For brevity we omitted the index 1 occurring in (42.4).

From fig. 1.27 it is clear that

$$\tan \theta_1 = \frac{v'_{c.m} \sin \theta_{c.m}}{v'_{c.m} \cos \theta_{c.m} + |\dot{\mathbf{R}}|}. \quad (43.28)$$

According to (42.4) and (42.7) we can write

$$\dot{\mathbf{R}} = \frac{m_1}{m_1 + m_2} \mathbf{v}_0, \quad \text{since } \mathbf{v}_1 = \mathbf{v}_0, \mathbf{v}_2 = 0,$$

$$\mathbf{v}'_{c.m} = \frac{\mu}{m_1} \mathbf{v}_0, \quad \text{since } \mathbf{v}'_{c.m} = \dot{\mathbf{v}}_0 - \dot{\mathbf{R}}.$$

Substituting into (43.28), we have

$$\tan \theta_1 = \frac{\sin \theta_{c.m}}{\cos \theta_{c.m} + \frac{m_1}{m_2}} \quad (43.29)$$

Formula (43.29) establishes* the relation between the angles $\theta_{c.m}$ and θ_1 .

It should be noted that, if the mass of the target particle is $m_2 \gg m_1$, then $\theta_1 \approx \theta_{c.m}$. The meaning of this result is obvious: a very heavy particle obtains no momentum from the incident particle and remains at rest — the centre-of-mass system coincides with the laboratory system.

Knowing the relation between the scattering angles in the laboratory and centre-of-mass systems, we can express the corresponding cross sections in terms of each other. Namely, writing the number of particles scattered into the solid angle $d\Omega_1$ and $d\Omega_{c.m}$ in the two coordinate systems

$$dN = 2\pi I_0 \sigma_1(\theta_1) \sin \theta_1 d\theta_1,$$

$$dN = 2\pi I_0 \sigma_{c.m}(\theta_{c.m}) \sin \theta_{c.m} d\theta_{c.m},$$

we find the relation sought for between the cross sections in the laboratory (σ_1) and centre-of-mass ($\sigma_{c.m}$) systems:

* It should be stressed that this refers only to elastic collisions, because in such collisions the relative velocity before and after the collision has the value v_0 .

$$\sigma_1(\theta_1) = \sigma_{c.m.}(\theta_{c.m.}) \frac{\sin \theta_{c.m.} d\theta_{c.m.}}{\sin \theta_1 d\theta_1}. \quad (43.30)$$

In contrast to the differential cross section, the total cross section is the same in all coordinate systems:

$$\sigma = 2\pi \int \sigma_1(\theta_1) \sin \theta_1 d\theta_1 = 2\pi \int \sigma_{c.m.}(\theta_{c.m.}) \sin \theta_{c.m.} d\theta_{c.m.}. \quad (43.31)$$

We shall also dwell on the calculation of the energy lost by the scattered particle. From fig. 1.27 we can write, on the basis of the cosine theorem, that

$$(v'_{c.m.})^2 = (v_1')^2 + \dot{R}^2 - 2\dot{R}(v_1') \cos \theta_1. \quad (43.32)$$

Substituting the values of \dot{R} and $v'_{c.m.}$ and denoting by E_1 and E'_1 the energy of the scattered particle respectively before and after collision in the laboratory system, we find

$$\frac{E'_1}{E_1} - \frac{2\mu}{m_2} \frac{\sqrt{E'_1}}{\sqrt{E_1}} \cos \theta_1 - \frac{m_2 - m_1}{m_2 + m_1} = 0. \quad (43.33)$$

Eq. (43.33) determines the dependence of E'_1 on the scattering angle θ_1 and the mass of the particles. The largest energy loss takes place when $m_1 = m_2$ and $\theta_1 = \frac{1}{2}\pi$ (in this case $\theta_{c.m.} = \pi$). In this case $E'_1 = 0$ and the energy is completely transferred to the target particle.

We shall not dwell on other details of the theory of scattering which are covered in the books cited.

Now we turn to the calculation of the radiation arising when a beam of charged particles passes by charges at rest. This phenomenon, called bremsstrahlung, gives rise to X-rays (with a continuous spectrum) and plays an important role in the deceleration of high-energy particles moving in matter.

Since it is just high-energy particles that are of basic interest, we shall restrict ourselves to this case.

When the incident particle has a sufficiently large energy its scattering angle is small, if the improbable processes of central collision are excluded. Considering the trajectory to be almost rectilinear, one can simplify the corresponding formulae.

Let us first of all consider the radiation emitted by one particle. Locating the origin at the scattering centre, one can write the following expression for the components of forces acting on the incident particle:

$$F_x = \frac{e_1 e_2}{r^3} x, \quad F_y = \frac{e_1 e_2}{r^3} y. \quad (43.34)$$

Since the motion takes place in the (xy) plane, the component F_z of the force is absent.

Assuming that the velocity of motion of the particle is constant and is not changed by the scattering, and that the deflection from the rectilinear trajectory is small *, one can put in the formulae for the force components

$$x \approx vt, \quad y \approx \rho, \quad dt \approx dx/v. \quad (43.35)$$

The energy emitted by the particle into the solid angle $d\Omega$ in passing by the scattering centre is found by integrating the formula (28.11) (in which it should be assumed that $m_2 \rightarrow \infty$) with respect to the transit time. Taking into account (43.35), we have

$$\begin{aligned} -\Delta E d\Omega &= \int dI dt = \frac{1}{4\pi c^3} d\Omega \int [\ddot{\mathbf{d}} \times \mathbf{n}]^2 dt = \\ &= \frac{1}{4\pi c^3} \frac{e_1^2}{vm_1^2} d\Omega \int_{-\infty}^{+\infty} [\mathbf{F} \times \mathbf{n}]^2 dx. \end{aligned}$$

Developing the expression $[\mathbf{F} \times \mathbf{n}]^2$, we have

$$\begin{aligned} [\mathbf{F} \times \mathbf{n}]^2 &= F^2 - (\mathbf{F} \cdot \mathbf{n})^2 = F_x^2 + F_y^2 - (F_x n_x + F_y n_y)^2 = \\ &= (1 - n_x^2) F_x^2 + (1 - n_y^2) F_y^2 - 2F_x F_y n_x n_y, \end{aligned}$$

where n_x and n_y are the components of the unit vector \mathbf{n} in the direction of the solid angle $d\Omega$.

We now have to integrate three integrals. The first of them is

$$\int_{-\infty}^{+\infty} F_x^2 dx = e_1^2 e_2^2 \int_{-\infty}^{+\infty} \frac{x^2}{r^6} dx = 2e_1^2 e_2^2 \int_0^{+\infty} \frac{x^2 dx}{(x^2 + \rho^2)^3}.$$

* L.D.Landau and E.M.Lifshitz, *Mechanics* (Pergamon Press, Oxford, 1960).

Putting $x/\rho = \tan \varphi$, we find

$$\begin{aligned}
 \int_{-\infty}^{+\infty} F_x^2 dx &= 2e_1^2 e_2^2 \int_0^{+\infty} \frac{x^2 dx}{(x^2 + \rho^2)^3} = \\
 &= 2 \frac{e_1^2 e_2^2}{\rho^3} \int_0^{\pi/2} \frac{\tan^2 \varphi}{(\tan^2 \varphi + 1)^3} \frac{d\varphi}{\cos^2 \varphi} = \\
 &= \frac{2e_1^2 e_2^2}{\rho^3} \int_0^{\pi/2} \sin^2 \varphi \cos^2 \varphi d\varphi = \\
 &= \frac{e_1^2 e_2^2}{2\rho^3} \int_0^{\pi/2} \sin^2 2\varphi d\varphi = \frac{\pi e_1^2 e_2^2}{8\rho^3}.
 \end{aligned}$$

The second and third integrals are calculated in an analogous way:

$$\int_{-\infty}^{\infty} F_y^2 dx = 2e_1^2 e_2^2 \rho^2 \int_0^{\infty} \frac{dx}{r^6} = \frac{3\pi e_1^2 e_2^2}{8\rho^3},$$

$$\int_{-\infty}^{\infty} F_x F_y dx = 0,$$

since $F_x F_y$ is an odd function of x . Finally,

$$\begin{aligned}
 -\Delta E d\Omega &= \frac{e_1^4 e_2^2}{32c^3 v m_1^2} \frac{1}{\rho^3} \{ (1 - n_x^2) + 3(1 - n_y^2) \} d\Omega = \\
 &= \frac{e_1^4 e_2^2}{32c^3 v m_1^2} \frac{1}{\rho^3} \{ 4 - \sin^2 \theta \cos^2 \psi - \\
 &\quad - 3 \sin^2 \theta \sin^2 \psi \} \sin \theta d\theta d\psi.
 \end{aligned} \tag{43.36}$$

The energy emitted is inversely proportional to the velocity and third power of the impact parameter ρ , and depends in a rather complicated way

on the angles. A simple calculation gives for the total irradiated energy

$$-\Delta E_{\text{tot}} = -\int \Delta E \, d\Omega = \frac{\pi}{3} \frac{e_1^4 e_2^2}{c^3 m_1^2 v} \frac{1}{\rho^3}. \quad (43.37)$$

In practice the calculation of the energy loss of a fast charged particle per unit path length in matter is of basic interest. The target centres deflecting the particle are nuclei with a charge $e_2 = Ze$, where Z is the atomic number of the element and e is the charge of an electron (scattering by electrons can be disregarded). In scattering on each nucleus the particle emits energy determined by formula (43.37). Multiplying (43.37) by the number of nuclei per unit length in a cylindrical shell of radius ρ , $\rho + d\rho$ and integrating over all impact parameters, we find for the energy loss

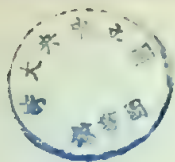
$$\begin{aligned} -\frac{dE}{dx} &= -\int \Delta E_{\text{tot}} N 2\pi\rho \, d\rho = \frac{2\pi^2 e_1^4 Z^2 e^2 N}{3c^3 m_1^2 v} \int_{\rho_{\min}}^{\infty} \frac{d\rho}{\rho^2} = \\ &= \frac{2\pi^2 Z^2 e^2 e_1^4 N}{3c^3 m_1^2 v} \frac{1}{\rho_{\min}}, \end{aligned} \quad (43.38)$$

where N is the number of nuclei per cm^3 and ρ_{\min} is a certain minimum value of the impact parameter.

If there were no minimum limiting distance, formula (43.38) would lead to a senseless result: the energy loss of the particle due to radiation along its path in matter would be infinitely large. However, in reality it turns out that the laws of classical physics become inapplicable at small distances from the nucleus. The calculation of the minimum value of the impact parameter can be carried out only on the basis of quantum mechanics (see also §17 of Part II).

PART II

THEORY OF RELATIVITY



General Principles of the Theory of Relativity

§ 1. The creation and significance of the theory of relativity

The development by the end of the 19th and the beginning of the 20th century of the theory of the electromagnetic field, and the improvement of experimental methods of investigating electromagnetic processes allowed investigators to carry out a persistent search for direct proofs of the existence of the hypothetical ether. Within the scope of this book we cannot describe the history of this search, which did not lead to the discovery of the ether but to the development of a new view of the physical world and to the complete rejection of the concepts of space and time which had been established in physics by the end of the last century.

These classical concepts were closely associated with the successes of classical mechanics. The principles of the classical view of physics can be expressed briefly in the following words:

1. A physical phenomenon can be considered to be thoroughly understood only when a mechanical model of it has been constructed.

2. The only possible form of physical law is a dynamical law of classical mechanics. As is known, in classical mechanics it is assumed that the specification of the forces which are acting and the initial conditions completely determines the motion of any mechanical system. Thus, the initial state determines completely the behaviour of the system at any subsequent instant. It is this statement which is contained in the idea of a dynamical law.

3. All physical processes take place in space and time. The properties of space and time are established in classical mechanics. Any physical theory must be constructed according to mechanics.

It was assumed that the properties of space reduce to:

- 1) the equivalence of all directions (isotropy);
- 2) the equivalence of all points of space (homogeneity);
- 3) the Euclidean nature of space.

It was assumed that, although the motion of physical bodies always takes place in space, the bodies in no way affect the properties of space.

In classical mechanics it was also considered possible to introduce a unique universal time flowing uniformly and equally, independent of the state of motion of physical bodies.

The creation of the theory of relativity by Einstein in 1905 led to a radical revision of ideas on the properties of space and time and the character of the electromagnetic field, and to the denial of the necessity and possibility of constructing mechanical models for all physical phenomena. The theory of relativity played a paramount role in the further development of contemporary physics, in particular, atomic and nuclear physics. This role consisted not only in making use of the important relations of the theory of relativity. The theory was the first to show that classical concepts, obtained from every-day experience, which seem to be obvious, turn out to be inadequate in going to new fields of investigation. Hence it can be correctly stated that the appearance of the theory of relativity signified the beginning of the development of a new, non-classical physics.

§2. Galilean transformations

In order to characterize the motion of bodies in space it is necessary to make use of a system of physical bodies between which an interaction, for example an electromagnetic interaction, exists. In addition it is necessary to make use of a method for the measurement of time. A method of measuring time, called a clock, is provided by any periodic process. Then, knowing the velocity of light and the time required for the light to travel from one body to another, one can determine the distances between the bodies. A set of bodies provided with clocks and located at distances determined in such a way is called a reference frame. Only when one makes use of a reference frame one can speak about a definite law of motion of a body in space. If the position of a body is referred to a reference frame at every instant of time,

then the set of all positions of the body in space forms a trajectory, and the sequence of the points of the trajectory represents a law of motion.

Any set of bodies, moving according to arbitrary laws, can be chosen as a reference frame. However, in the following we shall be interested in so-called inertial frames. By inertial reference frames we shall understand those frames in which Newton's law of inertia holds. In other words, in inertial reference frames the motion of bodies which are not acted upon by external forces is uniform and rectilinear. The special role of inertial reference frames is associated with the fact that in them the motion has its simplest form. In non-inertial reference frames, for example in a rotating reference frame, even the simplest rectilinear and uniform motion is described by very complicated relations.

Our problem is the comparison of the laws of motion of a body in different reference frames. If a certain physical law does not change in the transition from one reference frame to another, then we say that it is invariant under this transformation.

Long ago it had been established that mechanical phenomena are equivalent in all inertial reference frames. In other words, the laws of classical mechanics are invariant under transition from one inertial reference frame to others.

Let us consider two reference frames K and K' moving relative to each other. We shall call K' the moving frame, and K the frame at rest. The relative character of such a terminology will be particularly clear from what follows.

It is easy to obtain relations between the velocity and position of a moving body with respect to the two inertial reference frames. We direct the x -axis and x' -axis along the velocity vector \mathbf{v} of the relative motion. Then the relative motion will take place only along the positive direction of the x -axis. Moreover, we match the origins of the two frames at the initial instant $t = 0$ (fig. II.1).

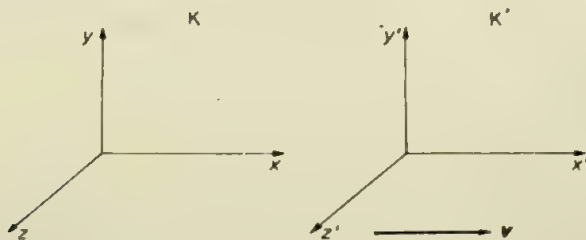


Fig. II.1

It should be noted, according to the above, that in order to find the laws of transformation from reference frame K to K' , the law of inertial motion of a given body must have the same form in the two reference frames. That is to say, in both frames the acceleration of such a body is the same and equal to zero, i.e.

$$\ddot{x} = \ddot{x}' = 0, \quad \ddot{y} = \ddot{y}' = 0, \quad \ddot{z} = \ddot{z}' = 0.$$

On integrating we find

$$\dot{x} = \dot{x}' + v, \quad \dot{y} = \dot{y}', \quad \dot{z} = \dot{z}'. \quad (2.1)$$

The second integration gives

$$x = x' + vt, \quad y = y', \quad z = z'. \quad (2.2)$$

Here we have tacitly assumed that time has an absolute character and is the same in all reference frames. For completeness we should then write

$$t' = t. \quad (2.3)$$

Formulae (2.2) and (2.3) are the Galilean law of transformation, while formula (2.1) is called the law of addition of velocities of classical mechanics. Of course, formulae (2.1) and (2.2) can easily be written also in vector form, without specifying the choice of orientation of the coordinate axes, as follows:

$$\dot{\mathbf{r}} = \dot{\mathbf{r}}' + \mathbf{v}, \quad \mathbf{r} = \mathbf{r}' + \mathbf{v}t. \quad (2.4)$$

The invariance of the laws of classical mechanics in the transition from one inertial reference frame to another is expressed mathematically by the fact that they are invariant under a Galilean transformation. This means that if in Newton's equations

$$m\ddot{x} = -\frac{\partial U}{\partial x}, \quad m\ddot{y} = -\frac{\partial U}{\partial y}, \quad m\ddot{z} = -\frac{\partial U}{\partial z},$$

one makes the substitution $x \rightarrow x', y \rightarrow y', z \rightarrow z'$, i.e. passes from the frame K to the frame K' , they will remain the same, if the law of transformation of coordinates and time is the Galilean law (2.2)–(2.3). Indeed, since the equa-

tions of motion involve only accelerations, under Galilean transformations we have

$$m\ddot{x}' = -\frac{\partial U}{\partial x'}, \quad m\ddot{y}' = -\frac{\partial U}{\partial y'}, \quad m\ddot{z}' = -\frac{\partial U}{\partial z'},$$

which is the same as the equations of motion in the reference frame K.

It should be emphasized that the reference frames K and K' are completely equivalent. We could consider the transition from the reference frame K' to K with equal success.

Thus, a uniform and rectilinear motion of the reference frame has no effect on mechanical processes taking place among a system of material points. This statement is called the Galilean principle of relativity. It should be noted that the term "the Galilean principle of relativity" was introduced in association with the creation of the theory of relativity. The term "relativity" emphasizes the complete equivalence of inertial reference frames. The terms "at rest" and "a uniform and rectilinear motion" are of relative character. In classical mechanics only a relative motion has meaning. On the other hand, the ideas of absolute rest and absolute motion have no real significance. The principle of relativity in mechanics is usually formulated by the words "the uniform and rectilinear motion of a system of material points has no effect upon the inner motion of the system". The principle of relativity in classical mechanics (the Galilean principle) is restricted to inertial reference frames.

The Galilean principle of relativity is based on the concepts of classical physics about space and time. This principle, as well as formula (2.4) for the addition of velocities which results from it, is confirmed by such a vast amount of experimental evidence, in particular that associated with the phenomena of the world around us, that it is adopted as being self-evident.

§3. Attempts to determine an absolute velocity

Shortly after the formulation of the Maxwell-Lorentz theory of the electromagnetic field the problem arose of its generalization to the case of moving bodies.

There is, however, a profound difference between the equations of classical mechanics and those of electrodynamics.

Maxwell's equations involve a characteristic velocity — the velocity of propagation of electromagnetic waves in a vacuum (the velocity of light). Hence

they are not invariant under Galilean transformations. One can easily verify this by the direct substitution of the velocity c by the sum $(c+v)$.

The question naturally arose as to the reference frame with respect to which the velocity of light is to be measured. The Lorentz classical electrodynamics appeared to give an unambiguous answer to this question: it is measured with respect to a hypothetical medium called the universal ether.

The ether was endowed with the properties of an all-pervading, homogeneous and isotropic medium, motionless and filling all space. In the Lorentz theory the existence of an absolute isolated reference frame was assumed. To move meant to move with respect to the ether, and the velocity of motion with respect to the ether was the absolute velocity.

Thus in the Lorentz theory, as distinct from classical mechanics, a decisive attempt to renounce the principle of relativity was made. The fact that the Maxwell-Lorentz equations, in contrast to the Newtonian ones, turned out not to be invariant under Galilean transformations, appeared to be a direct consequence of the renunciation of the principle of relativity.

It is clear that the basic problem confronting electrodynamics at the end of the 19th century was the problem of the experimental determination of an absolute velocity and of obtaining direct proof of the existence of the ether *.

We cannot here go into the history of the search for the ether, which can serve as an example of the inventiveness and persistence of many investigators.

We shall consider only the basic ideas of two possible experiments. Let a source and detector of electromagnetic waves be mounted on a body moving with a velocity v relative to the motionless ether. If the source-detector direction coincides with the direction of motion of the body with respect to the ether v , then light will traverse the source-detector distance l in a time $T_1 = l/(c-v)$. By measuring the time T_1 one can find the velocity v with respect to the ether. However, since c is very large, and the velocities attainable at the end of the last century were small, such a measurement was beyond the feasible accuracy of experiment. It was possible, however, to compare the time T_1 with the time T_2 during which light traverses the same

* See, for example, R. Becker, *Electromagnetic fields and interactions* (Blackie, London, 1964); W. Panofsky and M. Phillips, *Classical electricity and magnetism* (Addison-Wesley, 1964); and, in particular, L. I. Mandelshtam, *Sobr. soch. (Collected papers)*, Part V (Izdatelstvo Akademii Nauk SSSR, 1950). In these books the reader can acquaint himself with the history of the problem, as well as with the detailed methods of carrying out the experiments.

distance l in a direction perpendicular to the velocity v . During the time T_2 the detector traverses a path vT_2 with respect to the ether, so that the total path traversed by the light from the source to the detector is equal to $\sqrt{l^2 + v^2 T_2^2}$. Correspondingly, for the time T_2 we have

$$T_2 = \frac{\sqrt{l^2 + v^2 T_2^2}}{c},$$

or

$$T_2 = \frac{l/c}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

By making the rays travelling from the source to the detector along the direction of the velocity v interfere with those in the perpendicular direction, it was possible to determine the difference between T_1 and T_2 and thus the velocity v with a high degree of accuracy.

In 1881 such an experiment was carried out by Michelson, who made use of the velocity of the orbital motion of the Earth as the velocity of the source.

He made the rays which traversed the path from the source to the detector in the direction of motion of Earth interfere with those in the direction perpendicular to it. It should be noted that nowadays the accuracy of measurements by electronic techniques allows one to measure the difference $(T_1 - T_2)$ directly, without having recourse to interference.

To the surprise of Michelson's contemporaries, no difference between the times T_1 and T_2 could be detected. It turned out that T_1 and T_2 were equal to each other to a very high degree of accuracy.

The following direct experiment may serve as another basic experiment. Let a source of light be moving, and a detector be at rest with respect to the ether. Then the dependence of the velocity of light on the velocity of the source can be found directly. In 1912 De Sitter proposed choosing the radiation of the so-called binary stars as radiation coming to the Earth from the moving source. Binary stars represent two stars close to each other, rotating about a common centre of mass. By observing the velocity of light emitted as the star was moving towards the Earth and when it was moving in the opposite direction (after half of the period of revolution) one could determine the velocity of the star with respect to the ether. However, here also no effect of the motion of the source on the value of the velocity of light was discovered.

A number of attempts were made to explain the negative result of this and many other similar experiments (for example, a change in the law of interaction between charges with a change in the value of their absolute velocity; see §20). However, all these attempts turned out to be unsuccessful. The solution of the problem was given only in the Einstein theory of relativity.

§4. Postulates of the Einstein theory of relativity

The negative result of Michelson's experiment led Einstein to revise the basic concepts of classical physics and, above all, the notions on the properties of space and time.

As a result he created the theory of relativity, which is also called the special theory of relativity.

The theory is based on two principles or postulates:

- 1) Einstein's principle of relativity,
- 2) the principle of the existence of a limiting velocity of propagation of interactions.

According to Einstein's principle of relativity, a uniform and rectilinear motion of bodies has no effect on processes taking place in them. In other words, all laws of nature are the same in inertial reference frames. If in a certain inertial reference frame an arbitrary law of nature is expressed in the form of an equation in which a physical quantity is a function of coordinates and time, then, performing the transformation of the coordinates and time to another inertial reference frame, we must obtain the same functional dependence of the physical quantity on the new coordinates and time. This statement is briefly formulated by the words: "the laws of nature are invariant under a transformation from one inertial reference frame to another". It is clear that Einstein's principle of relativity is a generalization of the Galilean principle of relativity. The latter established the relativity of inertial motion and the impossibility of introducing the notions of absolute motion and absolute rest within the framework of classical mechanics. The negative result of Michelson's experiment, as was first realized by Einstein, pointed out that the ideas of absolute motion and rest have no meaning in the theory of the electromagnetic field.

However, there is a profound difference between the Galilean and Einstein's principle of relativity. In the latter the transformation from one inertial reference frame to another is not associated with formulae for the transformation of coordinates and the law of addition of velocities of classical mechanics. Indeed, as an example, Maxwell's equations do not satisfy these

transformations. Hence, in Einstein's theory, a new law of transformation of coordinates and time had to be found for going from one inertial reference frame to another. The second postulate of the theory of relativity stating that any interactions between bodies propagate in vacuum with a universal finite velocity equal to the velocity of light in vacuum, $c = 3 \times 10^{10}$ cm/sec, and independent of the motion and state of the bodies serves this purpose. It is obvious that this postulate expresses directly the result of Michelson's experiment.

The second postulate of the theory of relativity was closely associated with the development of electrodynamics. It demonstrated clearly the inadequacy of the theory of action at a distance of classical mechanics. In electrodynamics it was established that there exists a finite velocity of propagation of electromagnetic interactions, numerically equal to the velocity of light in vacuum. Theoretical studies carried out later on by Einstein in association with the development of the so-called general theory of relativity showed also that the gravitational interaction has the nature of waves propagating in vacuum with the velocity of light. It is beyond doubt that the specific interaction between nuclear particles has the character of a short-range action.

It cannot be excluded that further development of physics may lead to the discovery of new forms of interaction. However, the principle of the limiting velocity of propagation of interactions expresses the hypothesis that the velocity of propagation of interactions in vacuum has a universal character and is associated directly with the properties of space and time and not with the physical nature of the interactions.

The existence of the limiting velocity of propagation of interactions indicates that there is a certain connection between intervals of space and time. This connection will be demonstrated more clearly in analysing the conclusions of the theory of relativity: At the same time, the existence of the limiting velocity automatically assumes the restriction of the velocity of motion of material bodies to the value c . If any particles could move with a velocity higher than that of light, then these particles could bring about an interaction between bodies with a velocity higher than the limiting velocity. Thus, Einstein's second postulate restricts the value of all velocities of motion possible in nature and the velocity of propagation of interactions to the value c .

The principle of the existence of a limiting velocity of propagation of interactions is closely associated with Einstein's principle of relativity. Indeed, it is easily seen that, if the velocity of propagation of interactions depended on the velocity of particles or on the nature of the interaction (i.e. if it were different for the electromagnetic and gravitational interaction), then the

principle of relativity would be violated. For example, if the velocity of light depended on the velocity of a rectilinear and uniform motion of the source of light, then the latter could be determined experimentally.

The propagation of interactions is, in the theory of relativity, often called the propagation of signals. By a signal one understands any interaction between bodies at a finite distance from each other and in a state of rest or relative motion. The principle of the existence of a finite velocity of propagation of interaction is called the principle of existence of a finite velocity of propagation of signals. The overall content of the theory of relativity follows from its two postulates. In particular, formulae for the transformation of coordinates and time, in place of the Galilean transformation formulae, result directly from the basic postulates of Einstein. At present both postulates of the theory of relativity are confirmed by a whole set of experimental data obtained in investigating atomic and nuclear processes, motions of fast particles in accelerators, and other devices.

In what follows we shall present a number of examples illustrating the last statement.

§5. The Lorentz transformation

Proceeding from the postulates of Einstein's theory of relativity formulated above one can find the law of transformation relating space coordinates and time in two reference frames in uniform rectilinear motion relative to each other.

- Let x, y, z, t and x', y', z', t' be the coordinates and time in inertial reference frames K and K' , and v the velocity of their relative motion.

There are no grounds for assuming that the time in the frame K' coincides with the time in the frame K , as was implicitly assumed in classical physics.

To simplify the calculations we shall choose the direction of velocity to be the direction of the x -axis and x' -axis, as is shown in fig. II.1.

We assume that at a certain instant t' at the point with coordinates (x', y', z') there occurs a physical process, which we shall call for brevity an event. Our problem is to find the "coordinates" of this event in the reference frame K , i.e. to find the values (x, y, z, t) characterizing the same physical process in the reference frame K .

To establish the analytical relationship between the values (x, y, z, t) and (x', y', z', t') we consider the propagation of a spherical electromagnetic wave in the two reference frames.

We choose as the time origin, $t = 0$, the instant at which the origin of the

reference frame K' coincides with the origin of the reference frame K . Let the spherical electromagnetic wave begin to propagate at the instant $t = 0$ from the origin. In the frame K the equation of the wave front has the form

$$x^2 + y^2 + z^2 - c^2 t^2 = 0. \quad (5.1)$$

Since, according to Einstein's principle of relativity, the law and velocity of propagation of a wave must be the same in all inertial reference frames, in addition to (5.1) we can with equal justification write the equation of the spherical wave in the reference frame K' :

$$(x')^2 + (y')^2 + (z')^2 - c^2 (t')^2 = x^2 + y^2 + z^2 - c^2 t^2 = 0. \quad (5.2)$$

The formulae for the transformation of coordinates and time must: 1) not violate the relations (5.1) and (5.2), and 2) be linear. The requirement of linearity is associated with the homogeneity of space, which means that there are no points which stand out in particular for their properties.

First of all we note that, since the motion of the reference frame K' is along the x -axis, the transformation of the coordinates (y, z) must have the form

$$y' = y, \quad z' = z. \quad (5.3)$$

The law of transformation of x' in terms of x can be written from the following considerations: if at the instant $t = 0$ the origins of the reference frames K and K' coincided, then the coordinate of the point $x' = 0$ in the reference frame K is written as follows: $x = vt$. Consequently, in the most general case one can write that

$$x' = \alpha(v)(x - vt), \quad (5.4)$$

where the coefficient $\alpha(v)$ depends only on the velocity of the relative motion.

Making no arbitrary assumptions about the coincidence of time in the two reference frames, we can write t' in the form of a linear homogeneous function of x and t :

$$t' = \beta t + \gamma x. \quad (5.5)$$

The coefficients β and γ can, generally speaking, depend on the velocity v . If

it turned out that $\gamma = 0$ and $\beta = 1$, then we would come back to the Galilean transformation.

In order to determine the coefficients α , β and γ , corresponding to the requirements of Einstein's principle of relativity, we have to substitute (5.4) and (5.5) into (5.2). This gives

$$\alpha^2(x - vt)^2 + y^2 + z^2 - c^2(\beta t + \gamma x)^2 \equiv x^2 + y^2 + z^2 - c^2 t^2.$$

In order for this identity to be satisfied it is necessary to equate the coefficients of x^2 , t^2 and xt :

$$\alpha^2 - c^2 \gamma^2 = 1,$$

$$\alpha^2 v^2 - c^2 \beta^2 = -c^2,$$

$$\alpha^2 v + c^2 \beta \gamma = 0.$$

From these three equations we find the unknown values of α , β and γ :

$$\alpha = \beta = \frac{1}{\sqrt{1 - v^2/c^2}},$$

$$\gamma = -\frac{\alpha v}{c^2} = -\frac{v}{c^2 \sqrt{1 - v^2/c^2}}.$$

We have chosen the positive sign of the square root.

Substituting the values of α , β and γ into the formulae for the transformation of coordinates (5.4) and (5.5) and taking into account (5.3), we find

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \quad (5.6a)$$

$$y' = y, \quad (5.6b)$$

$$z' = z, \quad (5.6c)$$

$$t' = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}. \quad (5.6d)$$

Formulae (5.6a)–(5.6d) are called the Lorentz transformation formulae.

According to Einstein's principle of relativity these transformations replace the Galilean transformations.

Before going on to a discussion of the consequences of the Lorentz transformation we shall write formulae for the inverse transformation from the frame K' to K . On the basis of Einstein's principle of relativity the reference frames K and K' are completely equivalent. We could repeat all the previous reasoning, taking the initial frame to be K' and not K . However, in this case the velocity of the relative motion is equal to $(-v)$ and not to v . Hence we obtain:

$$x = \frac{x' + vt'}{\sqrt{1 - v^2/c^2}}, \quad (5.7a)$$

$$y = y', \quad (5.7b)$$

$$z = z', \quad (5.7c)$$

$$t = \frac{t' + vx'/c^2}{\sqrt{1 - v^2/c^2}}. \quad (5.7d)$$

The same result is found if the eqs. (5.6a)–(5.6d) are solved with respect to non-primed quantities.

The importance of the consequences following from the Lorentz transformation leads us to stress once more that their derivation is based only on Einstein's principle of relativity, the principle of the constancy of the limiting velocity of propagation of interactions, and the assumption of the uniformity of all points of space and time. These propositions are at present confirmed by a vast amount of experimental material and their validity is beyond any doubt.

A remarkable feature of the Lorentz transformation is the fact that, at relative velocities of motion small in comparison with the velocity of light, it goes over into the Galilean transformation. Indeed, at $v \ll c$ one can disregard values of the second order of smallness containing $(v/c)^2$ and write

$$x \approx x' + vt', \quad t' \approx t,$$

which are the same as formulae (2.2) and (2.3).

Thus, in the limiting case $v \ll c$ the transformation laws of the theory of relativity and classical mechanics are the same. This means that the theory of relativity does not reject the Galilean transformation as incorrect but includes

it in a valid law — the Lorentz transformation — as a particular case holding at velocities of motion which are small in comparison with the velocity of light.

In what follows we shall see that this reflects the general relation between the theory of relativity and classical physics. The laws and relations of the theory of relativity go over into the laws of classical physics in the limiting case of velocities small in comparison with the velocity of light c .

§6. Consequences of the Lorentz transformation.

Space and time intervals

The Lorentz transformation leads to conclusions radically contradicting the usual notions of the properties of space and time which arose from everyday experience and which we formulated in §2. Indeed, from the Lorentz transformation it follows directly that concepts of space and time intervals are relative. In other words, the ideas “the size of a body” or “the time lapse between two physical events” have no absolute character and are different in different reference frames.

Consider first of all the concept of spatial extension (length). Let there be a body at rest in a certain reference frame K' . We shall henceforth call this body the scale. The scale is not acted upon by any forces which could deform it and change its size. We denote by L_0 the length of the scale in the direction of motion (x' -axis). This length, measured in that reference frame in which the scale is at rest (frame K'), will be called the proper length of the scale. By means of a Lorentz transformation we shall find the length of the scale in the reference frame K , i.e. the length of the scale moving with a velocity v relative to the frame K . In the reference frame K' let the coordinates of the beginning and end of the scale be respectively x'_1 and x'_2 . We shall find these coordinates in the reference frame K .

Since the scale moves relative to the reference frame K , it is necessary, in order to measure its size, to fix the coordinates of its beginning and end at the same instant, measured in the reference frame K . For the realization of this measurement it would be possible to fix at an instant t the position of the beginning and end of the scale by means of a light signal coming from the reference frame K' .

At a certain instant t let the beginning and end of the scale have the coordinates x_1 and x_2 in the reference frame K . By means of formula (5.6a) we find

$$x'_2 = \frac{x_2 - vt}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad x'_1 = \frac{x_1 - vt}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

On subtracting we have

$$x'_2 - x'_1 = \frac{x_2 - x_1}{\sqrt{1 - \frac{v^2}{c^2}}},$$

or, denoting the difference between the coordinates of the beginning and end of the scale (the length of the scale in the reference frame K) by L , we obtain

$$L = L_0 \sqrt{1 - \frac{v^2}{c^2}}. \quad (6.1)$$

We see that the length of a scale moving with a velocity v relative to the reference frame K turns out to be $\sqrt{1 - v^2/c^2}$ times smaller than its proper length. This contraction of the size of a body is often called the Lorentz contraction. Since the dimensions of the scale in the direction perpendicular to the velocity remain unchanged, the volume of the scale turns out to be connected with its proper volume by the formula

$$V = V_0 \sqrt{1 - \frac{v^2}{c^2}}. \quad (6.2)$$

Thus, the length and volume of a scale which does not undergo the action of external forces turn out to be relative quantities. In other words, the statement that the distance between two points of space is equal to L has no meaning without specifying to which reference frame this quantity is referred. The distance between two points depends on the motion of the reference frame.

In classical physics the absolute character of the notion of the length of a scale was considered as something self-evident. Here lies the fundamental difference between views on the properties of space in the theory of relativity and classical physics.

It should be borne in mind that the two reference frames K and K' are completely equivalent. Hence, if the scale is at rest in the reference frame K, then its length in the reference frame K' will be smaller than in the frame K in the same ratio. There is a complete reciprocity between the two reference

frames. One can verify this by a direct calculation using the Lorentz transformation formulae (5.7a)–(5.7d).

It is easy to show that the negative result of Michelson's experiment follows automatically from the existence of the Lorentz contraction. Indeed, when light passes along the direction of motion of the Earth and in the opposite direction, from the point of view of an observer at rest (to which all the reasoning of §3 refers) the length l must be reduced $\sqrt{1-v^2/c^2}$ times. From the point of view of the observer at rest the time T_1 needed for a ray of light to traverse the total path is equal to

$$T_1 = \frac{l\sqrt{1-\frac{v^2}{c^2}}}{c+v} + \frac{l\sqrt{1-\frac{v^2}{c^2}}}{c-v} = \frac{2l}{c} \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}.$$

Correspondingly, the time difference is

$$T_2 - T_1 = 0$$

It should be stressed that the contraction of length — the compression of a body in the direction of motion — is of a purely kinematical character. No internal tensions causing a deformation arise in the body. In this sense one can speak about a "rigid" or, more precisely, non-deformable body in the theory of relativity. On the other hand, the notion of an absolutely rigid body is incompatible with the inferences of the theory of relativity. Indeed, if one assumes the existence of an absolutely rigid body, i.e. a body with invariable distances between all the particles constituting it, then such a body could be used for the transmission of an interaction with a velocity as large as one likes. A blow delivered to one of its ends would be transmitted to the other end with infinitely large velocity. Hence, from the point of view of the theory of relativity, the existence of absolutely rigid bodies cannot be assumed, even as an idealization.

The concept of time also undergoes a very fundamental modification in the theory of relativity.

At a point x' in the reference frame K' , let a certain physical process take place during a time interval $\Delta t_0 = t'_2 - t'_1$, where t'_1 and t'_2 are the time of the beginning and end of the process. Then in the reference frame K for the instants t_1 and t_2 we can write

$$t_2 = \frac{t'_2 + \frac{vx'}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad t_1 = \frac{t'_1 + \frac{vx'}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Subtracting we find the time interval from the beginning to the end of the process in the reference frame K:

$$\Delta t = t_2 - t_1 = \frac{\Delta t_0}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (6.3)$$

The time Δt_0 , measured in the reference frame moving together with the body in which the process takes place, is called the proper time.

Formula (6.3) shows that the proper time Δt_0 between two physical events is $\sqrt{1 - v^2/c^2}$ times smaller than the time lapse between these events in the reference frame K.

In the theory of relativity it is customary to speak about the comparison of the clock run in different inertial reference frames. The clock is understood to be an arbitrary periodical process. Then it can be said that the time shown by the clock depends on the velocity of its motion. A clock moving relative to a certain reference frame runs more slowly from the point of view of this frame than a clock at rest in it (the latter clock being identical with the moving one)

Thus, as distinct from Newtonian physics, the course of time turns out to be dependent on the state of motion. There is no universal time, and the notion of the time interval between two physical events turns out to be relative. Here it should again be stressed that there is a complete reciprocity between the reference frames K and K'. The above reasoning could be reversed. If a physical process takes place at the point x in the reference frame K and lasts Δt , then in the frame K' it will last

$$\Delta t' = \frac{t_2 - \frac{vx}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}} - \frac{t_1 - \frac{vx}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{\Delta t}{\sqrt{1 - \frac{v^2}{c^2}}},$$

as is seen from the Lorentz transformations (5.6a)–(5.6d).

Formula (6.3) for the change in rate of the moving clock has been experimentally confirmed in several ways. The most obvious of them is the follow-

ing. In cosmic rays the decay of the positive and negative muon (μ^+ - and μ^- -meson, with a mass of 207 electron masses) into a positron (electron) and two neutrinos is observed. The decay has been observed for μ -mesons decelerated almost to rest as well as when they are moving with a velocity close to that of light. The lifetimes of a moving meson and a meson at rest are connected by the relativistic relation

$$\tau_{\text{mov}} = \frac{\tau_{\text{rest}}}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Since v is close to the velocity of light, τ_{mov} must be considerably larger than τ_{rest} .

A number of experimental methods permit the determination of the value of τ_{rest} , which turns out to be 2×10^{-6} sec. If the lifetime of mesons were not dependent on the velocity, they would traverse a path equal to $v\tau_{\text{rest}} \approx 600$ m (for $v \approx c$). In reality, as shown by measurements, the mesons decay after traversing a path of about 20 km. To such a range there corresponds the lifetime

$$\tau_{\text{mov}} = \frac{20 \text{ km}}{c} \approx 7 \times 10^{-5} \text{ sec} \approx 50 \tau_{\text{rest}}$$

The relativistic change of the lifetime in this case turns out to be a very large effect.

Two questions naturally arise. First, why before the creation of the theory of relativity were all available experimental data in agreement with the Newtonian ideas of the absolute character of the length of the body and a unique universal time. Second, is the contraction of the length of moving bodies and the slowing down of the running of a moving clock real or apparent.

The answer to the first question is very simple. Before experiments designed to observe the motion of the Earth relative to the ether physicists did not encounter processes taking place with objects which moved with a velocity comparable to the velocity of light c . In other words, the velocities of all bodies observed in physics before the discovery of the electron were small in comparison with the velocity of light.

For velocities which are small compared with the velocity of light one can, with a sufficient degree of accuracy, make use of the earlier notions of space and time. Moreover, at the time of Einstein's creation of the theory of relativity Michelson's experiment was the sole undeniable indication of the inade-

quacy of classical physics. As will be seen from what follows, in the course of the past 50 years the situation has changed radically. The theory of relativity has become one of the bases of contemporary theoretical physics and a number of branches of experimental physics. Numerous experimental confirmations of the theory of relativity will, in part, be considered below. In particular, atomic and especially nuclear physics, as a rule, investigate the processes and behaviour of moving particles whose velocities are very close to the velocity of light. All the basic relations of the theory of relativity are widely used in nuclear physics for purely practical calculations. Some of these will be given below.

In considering the second question it should be emphasized that the rather widely used terms "an apparent contraction of the scale" and "an apparent change of the running of a clock" are inconvenient. Usually the authors wish to stress by the term "apparent" the purely kinematical character of the contraction. At the same time, the contraction of the scale and the slowing down of the running of the clock are a real and objective fact which is by no means connected with any illusions of the observer. It goes without saying that all values of the length of a given scale or time interval obtained in different reference frames are equivalent. They are all "correct". The difficulty in understanding these statements is associated solely with our habit of considering the notions of length and time interval as absolute notions, whereas in reality they are relative ones. Hence it is also senseless to ask which length of the scale is the true one and which is the apparent one, as it is senseless to say "in reality a given body is moving (or is at rest)". The concepts of length and time interval are as relative as are the concepts of motion and rest.

The true character of the contraction of a moving scale can be illustrated by the following example. Let there be two charged bodies A and B at rest in two separate reference frames. They are each spherical in the reference frame in which they are at rest. If one of the bodies, for example A, moves together with the reference frame K, then from the point of view of the reference frame K' its longitudinal (with respect to the direction of motion) dimension undergoes a contraction, whereas its transverse dimensions remain unchanged. The interaction between the bodies A and B then corresponds to the interaction between a charged sphere B and a charged ellipsoid A. In this case, from the point of view of the reference frame K', the body B is a sphere and the body A is an ellipsoid, whereas from the point of view of the reference frame K the body A remains a sphere and the body B transforms into an ellipsoid. However, the magnitude of the A-B interaction will be the same in the two reference frames.

An example of the calculation of the interaction between rapidly moving electric charges will be considered in §20.

§7. Einstein's law of addition of velocities and angular transformations

An important consequence of the Lorentz transformation is the relativistic law of addition of velocities, which in the theory of relativity replaces the Galilean law of addition of velocities.

It is found most simply by writing the Lorentz transformation formulae for the differentials of the space coordinates and time:

$$dx = \frac{dx' + v dt'}{\sqrt{1 - \frac{v^2}{c^2}}},$$

$$dy = dy',$$

$$dz = dz',$$

$$dt = \frac{dt' + \frac{v dx'}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Let x', y', z' be the coordinates of a material point moving in the reference frame K' . The components of the velocity of the point in the reference frame K' will be

$$u'_x = \frac{dx'}{dt'}, \quad u'_y = \frac{dy'}{dt'}, \quad u'_z = \frac{dz'}{dt'},$$

and in the reference frame K

$$u_x = \frac{dx}{dt}, \quad u_y = \frac{dy}{dt}, \quad u_z = \frac{dz}{dt}.$$

Dividing the differentials of the coordinates by the differential of time, we find

$$u_x = \frac{dx}{dt} = \frac{u'_x + v}{1 + \frac{u'_x v}{c^2}}, \quad (7.1)$$

$$u_y = \frac{dy}{dt} = \frac{u'_y \sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{u'_x v}{c^2}}, \quad (7.2)$$

$$u_z = \frac{dz}{dt} = \frac{u'_z \sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{u'_x v}{c^2}}. \quad (7.3)$$

Formulae (7.1)–(7.3) are called Einstein's law of addition of velocities. They replace the formulae of addition of velocities (2.1) of classical mechanics.

At $v \ll c$ Einstein's law of addition of velocities goes over directly into (2.1), since in this case

$$1 + \frac{u'_x v}{c^2} \approx 1, \quad \sqrt{1 - \frac{v^2}{c^2}} \approx 1.$$

As was to be expected, it follows from formulae (7.1) and (7.2) that the velocity of light c is the limiting velocity. If, for example, a particle in the reference frame K' moves along the x -axis with a velocity $u = u'_x = c$, then in motionless reference frame K its velocity is

$$u = \frac{c + v}{1 + \frac{cv}{c^2}} = c.$$

If a particle moves in the reference frame K' with a velocity smaller than the velocity of light, for example

$$u' = u'_x = c - \alpha \quad (\alpha > 0),$$

while the reference frame K' moves relative to K with a velocity $v = c - \beta$ ($\beta > 0$), then the velocity of the particle relative to the motionless frame K is equal to

$$u = \frac{(2c - \alpha - \beta)c}{2c - \alpha - \beta + \frac{\alpha\beta}{c}} < c.$$

Thus, the sum of two velocities, each of which is smaller than the velocity of light c , will always be smaller than the velocity of light. The sum of two velocities, one of which is equal to c and the other smaller than c , is equal to the velocity of light.

From the law of addition of velocities it follows directly that the value of an angle is relative and changes in transition from one inertial reference frame to the other. Since

$$\tan \theta = u_y / u_x,$$

where θ is the angle formed by the velocity vector of the particle and the x -axis, we find from (7.2) and (7.1) that

$$\tan \theta = \frac{u' \sqrt{1 - \frac{v^2}{c^2}} \sin \theta'}{v + u' \cos \theta'}, \quad (7.4)$$

where $u'_x = u' \cos \theta'$, $u'_y = u' \sin \theta'$.

The last formula expresses the law of angular transformations in the theory of relativity. It connects the angles θ' and θ , formed by the velocity vector with the x' -axis and x -axis respectively.

In conclusion it is emphasized that the velocity v should be a velocity with which a real body can move or a real process of interaction (signal) can propagate. One can imagine, without coming into conflict with the theory of relativity, processes having a velocity higher than the velocity c but which have a kinematical character and cannot transport bodies or realize interactions.

Consider, for example, the velocity of motion of an imaginary point P, at which a line A intersects a line B as the line B rotates. If the angle α at which A intersects B is arbitrarily small, and the length of the moving line is arbitrarily large, the velocity of motion of the point P can also be arbitrarily large. However, the motion of the imaginary point of intersection of the lines is not accompanied by a transfer of energy and cannot serve as a means for the transmission of signals and interactions.

§ 8. Simultaneity, short-range action and action at a distance

Let two physical events take place simultaneously at a time t' , at points x'_1 and x'_2 in an inertial reference frame K' . From the point of view of classical physics two events which are simultaneous in one reference frame take place simultaneously in all other inertial reference frames.

In the theory of relativity the situation is different.

Consider the inertial reference frame K relative to which the frame K' moves with a velocity v in the positive direction of the x -axis. In the reference frame K the first event takes place at an instant

$$t_1 = \frac{t' + \frac{x'_1 v}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

The second event takes place at an instant

$$t_2 = \frac{t' + \frac{x'_2 v}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Consequently, in the reference frame K the events do not take place simultaneously for there is a time lapse

$$\Delta t = t_2 - t_1 = \frac{v(x'_2 - x'_1)}{c^2 \sqrt{1 - \frac{v^2}{c^2}}}. \quad (8.1)$$

Moreover, depending on the sign of $(x'_2 - x'_1)$ the time interval Δt can be positive or negative, i.e. in the reference frame K "the first" event can take place earlier or later than "the second" one.

Thus, the concept of simultaneity turns out to be relative.

The sole, but very important exception is the case when two events occur at the same place, i.e. at an instant t' at a point x' . Then, according to (8.1), in all inertial reference frames (for any v) $\Delta t = 0$, i.e. the two events take place absolutely simultaneously.

The above illustrates particularly clearly that the theory of relativity is incompatible with the notion of action at a distance. Two events can be in a

mutual relationship as cause and effect only where they occur at the same place simultaneously as is required by the concept of short-range action. If, on the contrary, the cause and effect could be spatially separated (and the interaction propagated with an infinitely large velocity), then there would always exist an infinitely large number of inertial reference frames in which the effect would precede the cause.

It should be stressed that the concepts of relativity and simultaneity were assumed from the very beginning as the basis of the theory of relativity (in the form of the principle of a finite limiting velocity of propagation of interactions) and they can be considered only as an obvious example of the internal consistency of the theory.

§9. Absolute values in the theory of relativity. Intervals and proper time

The theory of relativity revolutionized the teaching of classical physics about the absolute character of space and time.

The relative character of space and time intervals seemed to be so paradoxical that the authors of a number of popular expositions of the theory of relativity, which appeared in particular around 1920, presented the ideological content of the theory by the scathing, but absolutely untrue aphorism: "The theory of relativity showed that all in the universe is relative".

In reality the situation is just the opposite. The task which the theory of relativity sets itself is to find absolute laws of nature which do not depend on the choice of the inertial reference frame *.

Thus, the theory of relativity by no means denies the existence of absolute quantities and concepts. It states only that a number of concepts, considered as absolute in classical physics, for example the magnitudes of space and time intervals, are in reality relative.

In this connection the opinion was often expressed that the term "theory of relativity" is unsuitable, since it does not reflect the content of this branch of physics. It was pointed out that the theory of relativity could with more justice be called "the theory of physical invariance". However, it should be borne in mind that at the time the theory of relativity was first introduced its critical aspect — the establishment of the relativity of space and time intervals — appeared to be more fundamental and new.

* In the general theory of relativity, which we cannot go into within the framework of this book, the problem of finding absolute laws of nature is extended to arbitrary reference frames.

The problem of finding absolute expressions for the laws of nature is closely connected with the finding of invariant, absolute quantities. The first of such quantities is the universal velocity of propagation of interactions — the velocity of light c . The second, also very important invariant quantity, is the so-called interval.

The notion of interval in the theory of relativity is a generalization of the usual notions of the interval (i.e. distance) between two points and the interval (i.e. lapse of time) between two events. Let there take place at a point in space with coordinates (x, y, z) at a time t , a physical phenomenon which we shall call an event. At another point x_1, y_1, z_1 at an instant t_1 another event takes place. Then the quantity

$$s = \sqrt{c^2(t_1 - t)^2 - (x_1 - x)^2 - (y_1 - y)^2 - (z_1 - z)^2} \quad (9.1)$$

is called the interval between the two events.

The invariance of the interval with respect to the Lorentz transformation can be checked by a direct calculation. In the moving reference frame K' we have

$$s' = \sqrt{c^2(t'_1 - t')^2 - (x'_1 - x')^2 - (y'_1 - y')^2 - (z'_1 - z')^2}$$

Obviously,

$$\begin{aligned} (x'_1 - x')^2 &= \frac{(x_1 - x)^2 + v^2(t_1 - t)^2 - 2v(x_1 - x)(t_1 - t)}{1 - v^2/c^2}, \\ c^2(t'_1 - t')^2 &= \frac{c^2(t_1 - t)^2 - 2v(x_1 - x)(t_1 - t) + \frac{v^2}{c^2}(x_1 - x)^2}{1 - v^2/c^2}, \\ (y'_1 - y')^2 &= (y_1 - y)^2, \\ (z'_1 - z')^2 &= (z_1 - z)^2. \end{aligned}$$

The substitution of these expressions into s' with some elementary calculations gives

$$s' = s.$$

Thus, the statement: "two physical events are separated by an interval s " is of absolute character. It is valid in all inertial reference frames.

The interval between two events taking place at infinitesimally close points with an infinitesimally small time separation is often considered. In this case the interval between two events is

$$ds = \sqrt{c^2 dt^2 - dx^2 - dy^2 - dz^2} . \quad (9.2)$$

The value of the interval s can be real or imaginary, depending on the sign of the expression under the radical.

Consider at first the case of a real interval:

$$c^2(\Delta t)^2 > (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 .$$

In this case it is always possible to find some reference frame in which the two events take place at the same point. For this it is necessary that the condition

$$\sqrt{c^2(\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2} = c \Delta t'$$

be fulfilled. In principle it can always be fulfilled for a positive value of the expression under the radical. Hence real intervals are called "time-like intervals". In particular, it is obvious that if two events take place involving the same physical system, the interval between these events is time-like. Indeed, during the time Δt between two successive events the system can traverse a path

$$\sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2} < c \Delta t ,$$

since its velocity is always smaller than the velocity of light. As an example of a time-like interval one can cite the interval between two events representing successive indications of the same clock.

An imaginary interval is called a "space-like interval". If two events are separated by a space-like interval, then one can always find a reference frame in which they take place at the same instant. For this it is necessary that the equality

$$\sqrt{c^2(\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2} = i \sqrt{(\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2}$$

be fulfilled. This equality always holds for negative values of the expression under the radical at the left.

We shall now return to the determination of the proper time and show that the proper time as well as the interval is an invariant, absolute quantity.

Let the inertial frame K' be given. At a certain point x', y', z' let there take place two successive events separated by a time interval dt_0 . It should be stressed that the time t_0 is measured by a clock which is at rest in the reference frame K' (or, as is said, by the proper clock of the reference frame K'), and that the time dt_0 is the proper time which has elapsed between the two events. The interval between the two events is by definition equal to

$$ds = \sqrt{c^2(dt_0)^2 - (dx')^2 - (dy')^2 - (dz')^2} = c dt_0.$$

Thus, the proper time is connected with the interval by the relation

$$dt_0 = c^{-1} ds \quad (9.3)$$

and is invariant.

The proper time can be expressed in terms of the time dt in an arbitrary reference frame, i.e. in terms of the time measured by a clock moving with respect to K' with a velocity $(-v)$, by substituting the expression for ds into (9.3):

$$\begin{aligned} dt_0 &= \frac{1}{c} \sqrt{c^2 dt^2 - dx^2 - dy^2 - dz^2} = \\ &= dt \sqrt{1 - \frac{1}{c^2} \left\{ \left(\frac{dx}{dt} \right)^2 + \left(\frac{dy}{dt} \right)^2 + \left(\frac{dz}{dt} \right)^2 \right\}} = \\ &= dt \sqrt{1 - \frac{v^2}{c^2}}. \end{aligned} \quad (9.4)$$

The finite interval of the proper time, t_0 , is equal to

$$t_0 = \int_0^{t_0} \sqrt{1 - \frac{v^2}{c^2}} dt. \quad (9.5)$$

It should be emphasized that formula (9.5) is derived for the case of the motion of the clock together with an inertial reference frame, i.e. a motion with a constant velocity.

Formula (9.5) is often applied to accelerated motion, considering v as a function of time. However, it should be borne in mind that an accelerated motion of a reference frame cannot be considered in the special theory of

relativity. Hence, in the case of accelerated motion, the quantity t_0 determined by formula (9.5) does not have the meaning of proper time but is a convenient quantity, invariant under Lorentz transformations.

§10. The invariance of physical laws under Lorentz transformations.

Four-dimensional formulation of the theory of relativity

According to the principle of relativity all physical laws — the laws of mechanics, electrodynamics, statistical physics and so on — must be the same in all inertial reference frames. This means that all physical laws must be formulated in such a way that they may remain invariant under Lorentz transformations. In the following the relations which are invariant under Lorentz transformations will be called relativistically invariant or Lorentz-invariant relations.

The equations of mechanics which are invariant under Galilean transformations obviously do not satisfy the requirement of invariance under Lorentz transformations and, consequently, their form must be altered. On the contrary, the laws of electrodynamics (i.e. Maxwell's equations, as will be shown later) were formulated from the very beginning in such a way that they turned out to be relativistically invariant. From the point of view of the theory of relativity, Maxwell's equations are a model of "a correctly formulated" physical law.

The realization of the general aims of the theory of relativity — i.e. finding relativistically invariant forms of physical laws — had a great effect on the overall further development of physics. Thus, for example, the progress made in the course of the last several decades in quantum mechanics and, in particular, in the quantum theory of fields, has been closely associated with the fulfilment of the requirements of the theory of relativity.

It should be noted that the requirement of the invariance of physical laws under certain transformations of reference frames is not a specific feature of the theory of relativity. It is well known that the requirement of invariance of physical laws under a rotation of the reference frame is associated with the isotropy of space.

Indeed, every physical law is formulated in such a way that the quantities involved in it refer to a certain system of coordinate axes. It is clear that the content of a physical law cannot depend on the orientation of the coordinate axes in space. For example, the Newtonian equations

$$m\ddot{x} = F_x \quad m\ddot{y} = F_y, \quad m\ddot{z} = F_z \quad (10.1)$$

do not depend on the orientation in space of the system of coordinate axes (x, y, z) to which the components of forces and accelerations refer. The equations of motion remain unchanged under any rotation of the axes of this reference frame: under rotation, every one of the components of the acceleration and force transforms according to one and the same law, so that the equalities (10.1) are not violated. This property of the invariance of physical laws under rotations of the reference frame can be formulated more precisely in the following way. In classical physics all physical laws are formulated in the form of equalities of the type

$$\alpha = \beta \quad (10.2)$$

or

$$\alpha = \beta, \quad (10.3)$$

$$\alpha_{ik} = \beta_{ik}. \quad (10.4)$$

The first of these contains the relation between scalar quantities, which remain unchanged under rotation of the coordinate axes. The second one relates vector quantities, which change under rotation of the coordinate axes. That is to say if for simplicity of notation one restricts oneself to the rotation about the z -axis through an angle φ , then the x -component and y -component of vectors change according to the well-known formulae of analytical geometry:

$$\begin{aligned} x &= x' \cos \varphi - y' \sin \varphi, \\ y &= x' \sin \varphi + y' \cos \varphi. \end{aligned} \quad (10.5)$$

Since, however, the components of any vectors, in particular those of the acceleration and force vectors, change according to this law, the equality (10.3) (or the equality (10.1) which is a particular case of it) is not violated.

The equality (10.4) shows in a general form that the tensor order of the quantities on the two sides of it are the same.

Thus, every physical law must be formulated in such a way that it will contain only quantities of the same tensor order. In classical mechanics the laws of transformation of coordinates which must leave physical laws unchanged reduce to the following:

1. Invariance under Galilean transformations.
2. Invariance under spatial translations and rotations of the coordinate frame.

3. Invariance under the replacement of t by $(t + t')$, expressing the uniformity of the flow of time.

4. Invariance under the change of the sign of time $t \rightarrow -t$, indicating the reversibility of the laws of mechanics, which are symmetric with respect to the future and past.

The theory of relativity puts forward, instead of condition 1, the more general requirement of the invariance of physical laws under Lorentz transformations. We shall often call this invariance relativistic invariance.

The conditions 2–4 are also preserved in the theory of relativity.

At first sight it might seem that the diversity of physical laws and the physical quantities occurring in them rule out a general method of approach to the establishment of their relativistically invariant formulations. In practice, however, this is not so.

In order to find a general method of working out relativistically invariant expressions we again turn to the expression for the interval ds .

We introduce a completely formal quantity

$$\tau = ict, \quad (10.6)$$

which will be called the fourth coordinate or the imaginary time. It goes without saying that τ , as an imaginary quantity, has no direct physical meaning.

By means of the imaginary time the interval can be written in the form

$$-ds^2 = dx^2 + dy^2 + dz^2 + d\tau^2 \quad (10.7)$$

In this notation Lorentz transformations take on a new interpretation.

We assume, without going into the physical content of our treatment for the present, that the quantities x, y, z, τ are orthogonal coordinates in a certain imaginary four-dimensional space.

The Lorentz transformation represents a linear transformation of the four coordinates (x, y, z, τ) which leaves the quantity ds^2 unchanged. It is easy to see that, from the geometrical point of view, the quantity ds^2 represents the square of the distance between two points in a four-dimensional space. Consequently, the Lorentz transformation is a linear transformation which leaves the distance between two arbitrary points in this space unchanged. From geometry it is known that there are only two such linear transformations: parallel displacement and rotation. Parallel displacement represents a trivial transformation which amounts to the change of the origin of the system of the coordinates x, y, z, t . Hence the only linear transformation leaving the

value of the interval unchanged is a rotation in the four-dimensional space (x, y, z, t) .

In what follows we shall confirm this conclusion by a direct calculation.

Such a geometrical interpretation of the Lorentz transformation, which is due to Minkowski, allows one to draw conclusions directly about the relativistically invariant form of physical laws.

Namely, in order that an expression may be relativistically invariant it must have the form

$$a = b, \quad (10.8)$$

where a and b are scalars, or

$$a_\alpha = b_\alpha, \quad (10.9)$$

where a_α and b_α are four-dimensional vectors having four components $(\alpha = x, y, z, \tau)$ and, in general,

$$a_{\alpha\beta\gamma\dots} = b_{\alpha\beta\gamma\dots}, \quad (10.10)$$

where $a_{\alpha\beta\gamma\dots}$ and $b_{\alpha\beta\gamma\dots}$ are four-dimensional tensors of an arbitrary rank. Under rotations of coordinate axes in the space (x, y, z, τ) all quantities involved in the relativistically invariant expression transform according to one and the same law, so that equalities of the type (10.8)–(10.10) are not violated. These conditions of invariance in the case of the Minkowski four-dimensional space represent a direct analogue of the conditions of invariance under a rotation of the reference frame in the real three-dimensional space.

It should be stressed that the introduction of the idea of a four-dimensional space with coordinates x, y, z, τ has a formal character. It is in no way equivalent to the statement of the existence of a real space with four dimensions.

The time coordinate τ is a purely imaginary quantity, which underlines its special character and basic difference from the space coordinates x, y, z . Nevertheless, the introduction of the time coordinate τ has a profound physical meaning. It points to the indissoluble connection between space and time, about which we have already spoken.

Later we shall have to reduce a number of important physical laws and relations to a relativistically invariant form. For this it is necessary to find their four-dimensional generalization. How this is to be done will be seen by concrete examples.

In the meantime we shall verify the fact that the rotation transformation in the four-dimensional space (x, y, z, τ) is identical with the Lorentz transformation. For simplicity of notation we assume, as before, that the motion of inertial reference frames is performed in the direction of matched axes x and x' . In a four-dimensional interpretation this corresponds to a rotation in the plane (x, τ) , the orientation of the axes y, z remaining unchanged.

If the rotation angle is denoted by φ , then by analogy with (10.5) one can write the relation between the initial coordinates (x, τ) and the transformed coordinates (x', τ') :

$$x = x' \cos \varphi - \tau' \sin \varphi, \quad (10.11)$$

$$\tau = \tau' \cos \varphi + x' \sin \varphi. \quad (10.12)$$

The rotation angle φ must, obviously, be different for different values of the velocity v . We write the transformations (10.11) and (10.12) for the τ -axis of the reference frame K' , i.e. for points which have $x' = 0$. Then, obviously,

$$x = -\tau' \sin \varphi,$$

$$\tau = \tau' \cos \varphi.$$

Dividing the upper expression by the lower one, we obtain

$$\tan \varphi = -x/\tau,$$

or

$$\tan \varphi = ix/ct = iv/c, \quad (10.13)$$

where v is the velocity of uniform motion of the origin of the coordinate frame K' (the point $x' = 0$) with respect to the reference frame K . From the equality (10.13) one can easily find the values of the quantities $\sin \varphi$ and $\cos \varphi$ involved in formulae (10.11) and (10.12):

$$\cos \varphi = \frac{1}{\sqrt{1 + \tan^2 \varphi}} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}},$$

$$\sin \varphi = \tan \varphi \cos \varphi = \frac{iv/c}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Hence

$$x = \frac{x' - i \frac{v}{c} \tau'}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (10.14)$$

$$\tau = \frac{\tau' + i \frac{v}{c} x'}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (10.15)$$

Passing over from τ to a time t , we see that formulae (10.14) and (10.15) are the same as the Lorentz transformation.

It is not out of place to stress the conditional character of the graphical representation of Lorentz transformations. The rotation angle in fig. II.2 is an imaginary angle. Of course, we cannot picture a rotation through an imaginary angle. The merits and shortcomings of the graphical representation of Lorentz transformations are clearly seen from the following. Let a clock be at rest at a certain point x' in the reference frame K' . This physical event at a moment τ'_1 is represented by the first point, and at a moment τ'_2 by the

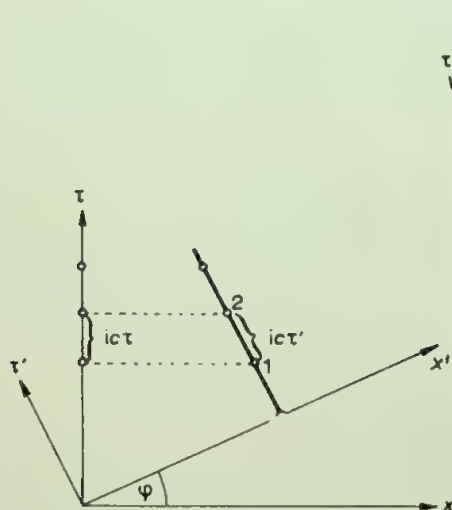


Fig. II.2

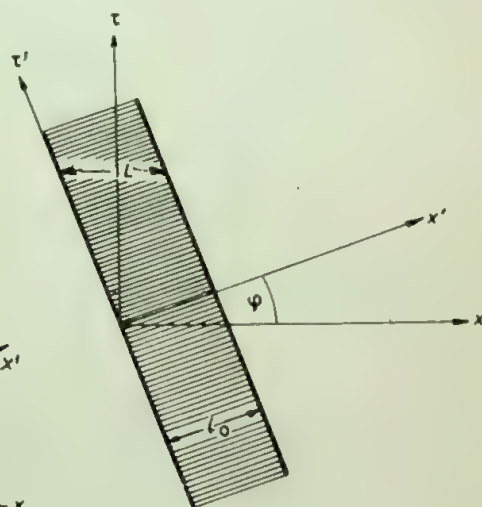


Fig. II.3

second point on the axis τ' . The time interval $\Delta\tau'$ is equal to the length of the section from point 1 to point 2. In passing over to the reference frame K (rotation through an angle φ) the segment $\Delta\tau'$ turns into a segment $\Delta\tau$ on the axis τ . We see clearly that it makes no sense to speak about which one of the reference frames is more correct and which one of the time intervals $\Delta\tau$ and $\Delta\tau'$ is the true time interval between the two physical events. A shortcoming of the geometrical consideration is the fact that the interrelation between $\Delta\tau$ and $\Delta\tau'$ in the drawing is the reverse of the true one: in the drawing $\Delta\tau$ is smaller than $\Delta\tau'$, whereas in reality it is larger in the ratio $1/\sqrt{1-v^2/c^2}$. The misrepresentation is due to the fact that we cannot show in the drawing the imaginary value of the angle φ and replace it by a real angle.

The length transformation (fig. II.3) and the Einstein theorem of addition of velocities allow an analogous geometrical interpretation. To the addition of velocities v_1 and v_2 there correspond two successive rotations in the plane (x, τ) : to the rotation through an angle φ_1 there corresponds the transition from the reference frame K to the reference frame K' moving with the velocity v_1 with respect to K; to the rotation through an angle φ_2 there corresponds the transition from the reference frame K' to a reference frame K'' moving with the velocity v_2 with respect to K'. Consequently, to the transition from K to K'', i.e. to the addition of velocities v_1 and v_2 there corresponds the rotation in the space (x, τ) through an angle

$$\varphi = \varphi_1 + \varphi_2.$$

Velocities v_1 and v_2 are connected with rotation angles φ_1 and φ_2 by the relation (10.13) which involves the tangent of the rotation angle. To an angle φ there corresponds a $\tan \varphi$ given by

$$\tan \varphi = \tan (\varphi_1 + \varphi_2) = \frac{\tan \varphi_1 + \tan \varphi_2}{1 - \tan \varphi_1 \tan \varphi_2}.$$

Substituting the values of $\tan \varphi$, $\tan \varphi_1$ and $\tan \varphi_2$ from (10.13) into the above relation, it is easy to arrive at the Einstein theorem of addition of velocities.

§ 11. Four-dimensional vectors and tensors.

Four-dimensional velocity and acceleration

We now pass on to four-dimensional vectors which, according to the results of the preceding paragraph, must figure in relativistically invariant expressions.

First of all we introduce the four-dimensional radius vector r_α ($\alpha = 1, 2, 3, 4$) whose projections onto mutually orthogonal coordinate axes are equal to x, y, z, τ .

In the Lorentz transformation – a rotation in the four-dimensional space – the components of the vector r_α transform according to the law

$$r_\alpha = \gamma_{\alpha\beta} r'_\beta,$$

such that the square of the vector r_α^2 remains invariant:

$$r_\alpha^2 = r_\alpha r_\alpha = x^2 + y^2 + z^2 + \tau^2 = \text{const}$$

For this the coefficients of the Lorentz transformation must obviously satisfy the requirement

$$\gamma_{\alpha\beta} \gamma_{\alpha\lambda} = \delta_{\beta\lambda},$$

often called the condition of orthogonality. Indeed, we have

$$r_\alpha^2 = (\gamma_{\alpha\beta} r'_\beta)(\gamma_{\alpha\lambda} r'_\lambda) = \gamma_{\alpha\beta} \gamma_{\alpha\lambda} r'_\beta r'_\lambda = \delta_{\beta\lambda} r'_\beta r'_\lambda = r'^2_\beta$$

This requirement imposes a necessary restriction upon the coefficients of the transformation $\gamma_{\alpha\beta}$. In the particular case where the Lorentz transformation corresponds to a rotation in the plane (x, τ) with unchanged values of y and z , making use of (10.14) and (10.15) one can easily write for $\gamma_{\alpha\beta}$

$$\gamma_{\alpha\beta} = \begin{vmatrix} \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} & 0 & 0 & -i\frac{v}{c} \\ \sqrt{1-\frac{v^2}{c^2}} & 0 & 0 & \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\frac{v}{c} & 0 & 0 & \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \\ \sqrt{1-\frac{v^2}{c^2}} & 0 & 0 & \sqrt{1-\frac{v^2}{c^2}} \end{vmatrix} \quad (11.1)$$

We generalize the definition of the vector r_α to the case of an arbitrary 4-vector. By definition, a set of quantities (components) a_x, a_y, a_z, a_τ , which under a Lorentz transformation (i.e. a rotation of the axes in the four-dimensional space) transform according to the same law as the components of the radius vector r_α , i.e. according to the law

$$a_\alpha = \gamma_{\alpha\beta} a'_\beta,$$

is called a 4-vector a_α . If one restricts oneself to rotations in the (x, τ) -plane, then, making use of the definition of $\gamma_{\alpha\beta}$ (or the analogy with (10.14) and (10.15)) one can write the transformation law in the form

$$a_x = \frac{a'_x - i \frac{v}{c} a'_\tau}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (11.2)$$

$$a_y = a'_y, \quad a_z = a'_z, \quad (11.3)$$

$$a_\tau = \frac{a'_\tau + i \frac{v}{c} a'_x}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (11.4)$$

For four-dimensional vectors, as for three-dimensional ones, one can introduce the idea of the scalar product

$$c = a_\alpha b_\alpha = \mathbf{a} \cdot \mathbf{b} + a_\tau b_\tau,$$

where c is a scalar quantity. The vectors a_α and b_α are called orthogonal if their scalar product is equal to zero. We do not consider other vectorial algebraic operations, since we shall not need them in what follows.

An important characteristic of a 4-vector is the scalar corresponding to the square of the vector $*$, i.e. to the scalar product

* Here and in what follows we make use of an abbreviated notation for summation. Repeated indices imply summation, for example

$$a_\alpha b_\alpha = \sum_\alpha a_\alpha b_\alpha, \quad a_\alpha^2 = a_\alpha a_\alpha = \sum_\alpha a_\alpha^2.$$

$$a_\alpha^2 = a_\alpha a_\alpha = a_x^2 + a_y^2 + a_z^2 + a_\tau^2 = \text{invar.} \quad (11.5)$$

The invariance of a_α^2 is seen directly from geometrical considerations: Lorentz transformations are a rotation in the four-dimensional space. The square of a 4-vector, like the square of the radius 4-vector r_α^2 , is not necessarily positive. If the square of the vector $a_\alpha^2 > 0$, then a_α is called a space-like vector. The vector for which $a_\alpha^2 < 0$ is called a time-like vector.

We shall consider the determination of two important 4-vectors: the 4-vector of velocity and the 4-vector of acceleration.

We have to find a 4-vector of velocity in the form of the derivative of the radius 4-vector with respect to a certain invariant — i.e. a scalar. The choice of this scalar is determined by the fact that, at small velocities $v \ll c$, the spatial components of the 4-vector of velocity must go over into the components of the ordinary velocity.

Because of this, it is natural to define the 4-vector of velocity by the relation

$$u_\alpha = \frac{dr_\alpha}{dt_0} \quad (11.6)$$

For the components of the 4-velocity we have, by eq. (9.4),

$$u_x = \frac{dx}{dt \sqrt{1 - \frac{v^2}{c^2}}} = \frac{v_x}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (11.7)$$

$$u_y = \frac{v_y}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (11.8)$$

$$u_z = \frac{v_z}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (11.9)$$

$$u_\tau = \frac{d\tau}{dt_0} = \frac{ic \, dt}{dt \sqrt{1 - \frac{v^2}{c^2}}} = \frac{ic}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (11.10)$$

For $v \ll c$ the three spatial components of the velocity are the same as the components of the usual three-dimensional velocity. The fourth component u_τ of the velocity is purely imaginary. Except for the factor (ic) it represents the coefficient of transition from the proper time dt_0 to the time dt .

An important feature of the 4-vector of velocity is the fact that its components are not independent of each other. Indeed, squaring the vector u_α , we find

$$u_\alpha^2 = u_x^2 + u_y^2 + u_z^2 + u_\tau^2 = -c^2. \quad (11.11)$$

Thus, the 4-vector of velocity is a time-like vector, and its absolute value is a constant. This property of the 4-vector is associated, of course, with the fact that the velocity of motion of material bodies cannot exceed the velocity of light.

Now we define the four-dimensional acceleration w_α as

$$w_\alpha = \frac{du_\alpha}{dt_0}. \quad (11.12)$$

Expressing w_α in terms of the velocity \mathbf{v} and the acceleration $\dot{\mathbf{v}}$, we find its components:

$$\begin{aligned} w_x &= \frac{du_x}{dt} \frac{dt}{dt_0} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \frac{d}{dt} \frac{v_x}{\sqrt{1 - \frac{v^2}{c^2}}} = \\ &= \frac{\dot{v}_x}{\left(1 - \frac{v^2}{c^2}\right)} + \frac{v_x(\mathbf{v} \cdot \dot{\mathbf{v}})}{c^2 \left(1 - \frac{v^2}{c^2}\right)^2}, \end{aligned} \quad (11.13)$$

$$w_y = \frac{\dot{v}_y}{\left(1 - \frac{v^2}{c^2}\right)} + \frac{v_y(\mathbf{v} \cdot \dot{\mathbf{v}})}{c^2 \left(1 - \frac{v^2}{c^2}\right)^2}, \quad (11.14)$$

$$w_z = \frac{\dot{v}_z}{\left(1 - \frac{v^2}{c^2}\right)} + \frac{v_z(\mathbf{v} \cdot \dot{\mathbf{v}})}{c^2 \left(1 - \frac{v^2}{c^2}\right)^2}, \quad (11.15)$$

$$w_\tau = \frac{dt}{dt_0} \frac{d}{dt} \left[\frac{ic}{\sqrt{1 - \frac{v^2}{c^2}}} \right] = \frac{i}{c} \frac{\mathbf{v} \cdot \dot{\mathbf{v}}}{\left(1 - \frac{v^2}{c^2}\right)^2}. \quad (11.16)$$

A simple calculation shows that the square of the four-dimensional acceleration is equal to

$$w_{\alpha}^2 = \frac{\dot{v}^2 - \frac{1}{c^2} [\dot{v}^2 v^2 - (\mathbf{v} \cdot \dot{\mathbf{v}})^2]}{\left(1 - \frac{v^2}{c^2}\right)^3} = \frac{\dot{v}^2 - \left[\frac{\mathbf{v}}{c} \times \dot{\mathbf{v}}\right]^2}{\left(1 - \frac{v^2}{c^2}\right)^3} > 0. \quad (11.17)$$

Thus the 4-acceleration is a space-like vector.

Differentiating the equality (11.11) with respect to t_0 , we find

$$u_{\alpha} \frac{du_{\alpha}}{dt_0} = u_{\alpha} w_{\alpha} = 0. \quad (11.18)$$

This last equality means that the vectors u_{α} and w_{α} are orthogonal to each other in the four-dimensional space.

In addition to the definition of 4-vectors one can introduce 4-tensors. The set of quantities $A_{\alpha\beta}$ which transform as the product of two vectors $a_{\alpha}b_{\beta}$, i.e.

$$A_{\alpha\beta} = \gamma_{\alpha\lambda} \gamma_{\beta\mu} A'_{\lambda\mu}, \quad (11.19)$$

is called a second-rank 4-tensor. The tensor $A_{\alpha\beta}$ represents a set of sixteen quantities (a second-rank tensor in the three-dimensional space represents a set of nine quantities). As in the case of a 4-vector, the τ -components of a 4-tensor, i.e. the quantities $A_{\alpha\tau}$, are purely imaginary, whereas all remaining components are real. Any tensor $A_{\alpha\beta}$ can be resolved into a symmetric part and an antisymmetric part (with respect to the permutation of subscripts). Writing

$$A_{\alpha\beta} = \frac{1}{2} (A_{\alpha\beta} + A_{\beta\alpha}) + \frac{1}{2} (A_{\alpha\beta} - A_{\beta\alpha}) = A_{\alpha\beta}^s + A_{\alpha\beta}^{as},$$

it is easily seen that

$$A_{\alpha\beta}^s = A_{\beta\alpha}^s,$$

$$A_{\alpha\beta}^{as} = -A_{\beta\alpha}^{as}.$$

Lorentz transformations leave this decomposition unchanged.

Purely antisymmetric tensors will play an important role in what follows. From the condition of antisymmetry it follows that such tensors have six independent components and can be written in the form of a table:

$$A_{\alpha\beta}^{\text{as}} = \begin{vmatrix} 0 & A_{xy} & A_{xz} & A_{x\tau} \\ -A_{xy} & 0 & A_{yz} & A_{y\tau} \\ -A_{xz} & -A_{yz} & 0 & A_{z\tau} \\ -A_{x\tau} & -A_{y\tau} & -A_{z\tau} & 0 \end{vmatrix} \quad (11.20)$$

Making use of the properties of the quantities $\gamma_{\alpha\beta}$ (11.1) one can find the law of transformation of the components of a 4-tensor. We restrict ourselves to the case of an antisymmetric tensor and write one of the components in detail:

$$\begin{aligned} A_{xy} &= \gamma_{x\lambda} \gamma_{y\mu} A'_{\lambda\mu} = \gamma_{xx} \gamma_{yy} A'_{xy} + \gamma_{x\tau} \gamma_{y\tau} A'_{\tau y} = \\ &= \frac{A'_{xy} - i \frac{v}{c} A'_{y\tau}}{\sqrt{1 - \frac{v^2}{c^2}}} \end{aligned}$$

The other terms reduce to zero by virtue of the definition of $\gamma_{\alpha\beta}$ and $A_{\alpha\beta}^{\text{as}}$. The law of transformation of the remaining components is easily found in the same way. As a result, for the transformation of the components of the antisymmetric tensor we obtain

$$\begin{aligned} A_{xy} &= \frac{A'_{xy} - i \frac{v}{c} A'_{y\tau}}{\sqrt{1 - \frac{v^2}{c^2}}}, & A_{x\tau} &= A'_{x\tau}, \\ A_{xz} &= \frac{A'_{xz} - i \frac{v}{c} A'_{z\tau}}{\sqrt{1 - \frac{v^2}{c^2}}}, & A_{y\tau} &= \frac{A'_{y\tau} + i \frac{v}{c} A'_{xy}}{\sqrt{1 - \frac{v^2}{c^2}}}, \\ A_{yz} &= A'_{yz}, & A_{z\tau} &= \frac{A'_{z\tau} + i \frac{v}{c} A'_{xz}}{\sqrt{1 - \frac{v^2}{c^2}}}. \end{aligned} \quad (11.21)$$

The invariants of 4-tensors, i.e. combinations remaining invariant under Lorentz transformations, are an important characteristic of 4-tensors. For anti-symmetric tensors of the second rank, which we shall need in what follows, the invariants are scalar quantities:

- 1) the product $A_{\alpha\beta}A_{\alpha\beta}$ (quadratic invariant),
- 2) the product $A_{\alpha\lambda}A_{\lambda\mu}A_{\mu\alpha}$ (cubic invariant) ,
- 3) the product $A_{\alpha\lambda}A_{\lambda\mu}A_{\mu\nu}A_{\nu\alpha}$ (biquadratic invariant).

One can verify the invariance of these quantities by the direct substitution of the transformation law (11.19) and the values of the quantities $\gamma_{\alpha\beta}$. Other scalar relations can either be expressed in terms of these three invariants or are equal to zero. Thus, for example, the scalar $A_{\alpha\alpha}$, representing the sum of diagonal elements, is equal to zero for antisymmetric tensors.

Relativistic Mechanics

§12. The dynamical equations of a material point

We now turn to the consideration of the dynamics of a material point in the theory of relativity.

As in classical mechanics, a material point is understood to be a body whose dimensions can be neglected. Having physical applications in mind, we shall often speak not about the motion of a material point but about the motion of a particle.

First of all we note that the Newtonian law of inertia is invariant under Lorentz transformations. Indeed, if a particle moves without acceleration in an inertial reference frame K , then under a linear transformation of the coordinates to another reference frame K' the motion will remain without acceleration. However, the equations of dynamics, which are invariant under Galilean transformations, possess no invariance properties under Lorentz transformations.

In order to find the relativistically-invariant form of the equations of dynamics it is necessary to represent them in the form of a four-dimensional relation of the type of (10.9).

The inertial properties of a body or a particle may be characterized by a certain scalar — the invariant mass or the rest mass m . The value of the rest mass is a constant characteristic of every kind of elementary particle. The

4-momentum p_α of a particle is defined as follows:

$$p_\alpha = mu_\alpha. \quad (12.1)$$

For the components we have:

$$\begin{aligned} p_x &= \frac{mv_x}{\sqrt{1 - \frac{v^2}{c^2}}}, & p_y &= \frac{mv_y}{\sqrt{1 - \frac{v^2}{c^2}}}, \\ p_z &= \frac{mv_z}{\sqrt{1 - \frac{v^2}{c^2}}}, & p_\tau &= \frac{icm}{\sqrt{1 - \frac{v^2}{c^2}}} \end{aligned} \quad (12.2)$$

In the limiting case $v \ll c$ the three spatial components of the momentum go over into the ordinary components of the particle momentum:

$$p_x = mv_x, \quad p_y = mv_y, \quad p_z = mv_z.$$

A natural relativistic generalization of the equations of Newtonian dynamics are the equations

$$\frac{dp_\alpha}{dt_0} = \frac{d}{dt_0} mu_\alpha = F_\alpha, \quad (12.3)$$

where F_α is a four-dimensional vector called the four-dimensional force or Minkowski force, and α runs through the values x, y, z, τ . The relativistically-invariant character of the eqs. (12.3) follows directly from what was said in §11: the right-hand and left-hand side of (12.3) involve four-dimensional vectors, which change in the same way under a four-dimensional rotation (Lorentz transformation), and the scalar m which does not change at all.

In the following we shall call the relations (12.3) the equations of relativistic dynamics.

We write the components of these equations. We have:

$$\frac{dp_x}{dt_0} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \frac{d}{dt} \frac{mv_x}{\sqrt{1 - \frac{v^2}{c^2}}} = F_x,$$

or

$$\frac{d}{dt} \frac{mv_x}{\sqrt{1 - \frac{v^2}{c^2}}} = F_x \sqrt{1 - \frac{v^2}{c^2}}. \quad (12.4)$$

For $v \ll c$ eq. (12.4) must go over into the ordinary Newtonian equation.

The left-hand side of formula (12.4) is the derivative of the momentum with respect to ordinary time. We require that the right-hand side of eq. (12.4) should be the component \mathcal{F}_x of the ordinary force. Consequently, the component F_x of the 4-force is connected with that of the ordinary force \mathcal{F}_x of classical mechanics by the relation

$$F_x \sqrt{1 - \frac{v^2}{c^2}} = \mathcal{F}_x. \quad (12.5)$$

Then formula (12.4) is written in the form

$$\frac{d}{dt} \frac{mv_x}{\sqrt{1 - \frac{v^2}{c^2}}} = \mathcal{F}_x. \quad (12.6)$$

For $v \ll c$ formula (12.6) reduces to the Newtonian equation.

Analogous relations can be written for the two remaining spatial components:

$$\frac{d}{dt} \frac{mv_y}{\sqrt{1 - \frac{v^2}{c^2}}} = \mathcal{F}_y, \quad (12.7)$$

$$\frac{d}{dt} \frac{mv_z}{\sqrt{1 - \frac{v^2}{c^2}}} = \mathcal{F}_z. \quad (12.8)$$

Now we write the fourth component of the relation (12.3). Making use of (11.10), we find

$$\frac{d}{dt} \frac{icm}{\sqrt{1 - \frac{v^2}{c^2}}} = F_\tau \sqrt{1 - \frac{v^2}{c^2}}. \quad (12.9)$$

To find the physical meaning of the component F_τ of the 4-force we multiply (12.3) by u_α and sum over all components α ($\alpha = x, y, z, \tau$).

Obviously, by virtue of (11.18) we have

$$u_\alpha \frac{dm u_\alpha}{dt_0} = F_\alpha u_\alpha = 0,$$

or

$$F_x u_x + F_y u_y + F_z u_z + F_\tau u_\tau = 0.$$

Substituting the values of $F_x, F_y, F_z, u_x, u_y, u_z$ and u_τ , we have

$$\begin{aligned} \mathcal{F}_x \frac{v_x}{\left(\sqrt{1 - \frac{v^2}{c^2}}\right)^2} + \mathcal{F}_y \frac{v_y}{\left(\sqrt{1 - \frac{v^2}{c^2}}\right)^2} + \\ + \mathcal{F}_z \frac{v_z}{\left(\sqrt{1 - \frac{v^2}{c^2}}\right)^2} + F_\tau \frac{ic}{\sqrt{1 - \frac{v^2}{c^2}}} = 0. \end{aligned}$$

From which it follows that

$$F_\tau \sqrt{1 - \frac{v^2}{c^2}} = \frac{i}{c} \mathcal{F} \cdot \mathbf{v}.$$

The right-hand side of this equation contains the work done by the force \mathcal{F} on the particle per unit time. Thus, the component F_τ of the 4-force turns out to be connected with the work done by the three-dimensional force \mathcal{F} :

$$F_\tau = \frac{i}{c} \frac{\mathcal{F} \cdot \mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (12.10)$$

Making use of (12.10) we write (12.9) in the form

$$\frac{d}{dt} \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} = \mathcal{F} \cdot \mathbf{v}. \quad (12.11)$$

On the right-hand side of (12.11) the product $\mathcal{F} \cdot \mathbf{v}$ gives the work done by the force on the particle in unit time. Consequently, on the left-hand side of the equation we must have the change of the energy in unit time.

Thus, we define the total energy of a particle as follows:

$$E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (12.12)$$

Finally, we find the expression for the acceleration. The equations of motion can be written in the form

$$\frac{d}{dt} \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{m}{\sqrt{1 - \frac{v^2}{c^2}}} \frac{d\mathbf{v}}{dt} + \frac{\mathbf{v}}{c^2} \frac{d}{dt} \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} = \mathcal{F},$$

or

$$\frac{m}{\sqrt{1 - \frac{v^2}{c^2}}} \frac{d\mathbf{v}}{dt} + \frac{\mathbf{v}}{c^2} \frac{dE}{dt} = \mathcal{F}. \quad (12.13)$$

By means of (12.11) one can rewrite (12.13) in the form

$$\mathbf{w} = \frac{d\mathbf{v}}{dt} = \frac{\sqrt{1 - \frac{v^2}{c^2}}}{\gamma} \left(\mathcal{F} - \frac{\mathbf{v}}{c^2} (\mathcal{F} \cdot \mathbf{v}) \right) \quad (12.14)$$

§ 13. Momentum, energy and mass in relativistic mechanics

We shall now discuss the properties of the mechanical quantities introduced in the preceding paragraph, i.e. the rest mass, momentum and energy.

According to (12.2) the connection between the rest mass m and the three-dimensional momentum \mathbf{p} is determined by the relation

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (13.1)$$

which for $v \ll c$ is the same as the usual expression of classical mechanics. Hence one would think that it could be taken that the scalar m is the same as the mass of a body moving with a small velocity.

However, as will be seen from what follows, the properties of the invariant mass m (the rest mass) differ essentially from those which are attributed to mass in classical mechanics. Namely, the rest mass does not satisfy the conservation law. There are physical processes in which the rest mass of the particles before the beginning of the process is not equal to that of the particles remaining after the end of the process.

Examples of phenomena of this kind will be given below. The singularity (from the point of view of the concept of mass in classical mechanics) of the rest mass m and the possibility of its non-conservation are particularly clear from the fact that the presence of rest mass is not a necessary property of particles.

For example, the existence in nature of elementary particles whose rest mass is equal to zero is beyond any doubt. Light quanta (photons) are such particles. According to all available experimental (see § 16) and theoretical data, the neutrino is also a particle with zero rest mass. The neutrino is a neutral particle playing an important role in nuclear processes (it arises in particular in β -decay).

It is clear that, if in the processes of mutual transformation, which play a most important role in the world of elementary particles, there are transitions of particles with non-zero rest mass into particles with zero rest mass, then the rest mass is not conserved. Such processes indeed occur in nature, and some of them are well investigated. Examples will be presented in § 16.

In spite of its unusual properties, the rest mass is a very important characteristic of bodies.

Every elementary particle has a well defined value of rest mass (including also the value zero), which does not change from sample to sample. Hence the rest mass is a fundamental characteristic of an elementary particle.

One can speak of the rest mass of a body consisting of many elementary particles in the same way as of the rest mass of an elementary particle. If the dimensions of the body are neglected, it can be considered as a material point with a rest mass m . The question as to how the rest mass of the body is connected with the rest masses of the constituent particles will be discussed below.

In addition to the rest mass, a mass $m(v)$, called the relativistic mass or simply mass and defined as the factor of proportionality between the vectors \mathbf{p} and \mathbf{v} , is often introduced:

$$\mathbf{p} = m(v) \mathbf{v}, \quad (13.2)$$

where

$$m(v) = \frac{m}{\sqrt{1 - \frac{v^2}{c^2}}} . \quad (13.3)$$

The relativistic mass depends on the velocity and is hence a function not only of the properties of the particle but also of the state of its motion.

However, it should be stressed that the relativistic mass $m(v)$ is not a relativistically-invariant quantity. Indeed, the quantity $\sqrt{1 - v^2/c^2}$ is not invariant (the law of transformation of $\sqrt{1 - v^2/c^2}$ is found most simply from (9.4), taking into account that dt_0 is an invariant and that the law of transformation of the time dt is known).

If a particle moves with different velocities relative to two reference frames, then its mass, measured by devices located in these reference frames, will be different.

The energy of a particle was determined by the relation (12.12). Before going on to a discussion of this relation, we note that the following important relation, connecting the time component of the momentum 4-vector and the energy, results from (12.12) and (12.2):

$$p_\tau = i E/c . \quad (13.4)$$

Thus, p_τ is the same as the energy of the particle apart from the constant factor. The importance of (13.4) lies in the fact that by means of it the momentum and energy may be combined into one 4-vector, called the energy-momentum 4-vector,

$$p_\alpha = (p_x, p_y, p_z, i E/c) . \quad (13.5)$$

The components of the energy-momentum 4-vector are not relativistically-invariant quantities. The three-dimensional momentum as well as the energy turn out to be relative quantities. In transition from one inertial reference frame to another the components of the energy-momentum 4-vector transform according to formulae (11.1)–(11.4), i.e.

$$p_x = \frac{p'_x + \frac{E'}{c^2} v}{\sqrt{1 - \frac{v^2}{c^2}}} , \quad (13.6)$$

$$p_y = p'_y, \quad (13.7)$$

$$p_z = p'_z, \quad (13.8)$$

$$E = \frac{p_x c}{i} = \frac{E' + v p'_x}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (13.9)$$

These relations show that under Lorentz transformations the energy and the components of the momentum are expressed in terms of each other. The energy and momentum are not separately invariant quantities, but the square of the 4-vectors, i.e. the quantity

$$p_\alpha^2 = p_x^2 + p_y^2 + p_z^2 + p_\tau^2 = p_x^2 + p_y^2 + p_z^2 - E^2/c^2 = \text{invar}. \quad (13.10)$$

Substituting the values of the energy (12.12) and the components of the momentum (12.2) into (13.10), one can easily calculate the value of this invariant, which turns out to be

$$I = p_x^2 + p_y^2 + p_z^2 - E^2/c^2 = -m^2 c^2. \quad (13.11)$$

Thus, the energy-momentum 4-vector is a time-like vector.

We now return to the determination of the energy (12.12) of a particle. For $v \ll c$ formula (12.12) goes over into

$$E = mc^2 \left(1 + \frac{v^2}{2c^2} + \dots \right) \approx mc^2 + \frac{1}{2}mv^2. \quad (13.12)$$

The second term is the same as the kinetic energy of a particle in classical mechanics. However, for $v = 0$ the energy of the particle outside a force field

$$E_0 = mc^2 \quad (13.13)$$

turns out to be different from zero.

At first sight it may seem that such a determination of the energy is arbitrary. Since the energy is found from the differential relation (12.11), it can be chosen to be

$$E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} + \text{const}. \quad (13.14)$$

If the arbitrary constant is chosen to be equal to $(-mc^2)$, then the energy determined in such a way at $v \ll c$ will be the same as the energy of the particle in classical mechanics.

In fact, however, it is easy to show that the constant must be set equal to zero, as was done in (12.12). We cannot require in advance that, without exception, all quantities of relativistic mechanics should assume the classical form for $v \ll c$. However, in any case it is beyond any doubt that for $v \ll c$ the Lorentz transformations must be the same as the Galilean transformations and, consequently, the ordinary law of addition of velocities must hold. In order that the transformation (13.6) of the momentum for $v \ll c$ may go over into the theorem of addition of velocities of classical mechanics, the condition $E' \rightarrow mc^2$ must be fulfilled. Then from (13.6) it follows that

$$p_x = p'_x + mv,$$

or

$$v_x = v'_x + v.$$

If the energy were determined by formula (13.14) and not by (12.12) and if for $v \ll c$ it tended to the limit $E' \rightarrow 0$, then the Lorentz transformation for the velocity would not reduce to the formula of addition of velocities of classical mechanics.

Thus, the theory of relativity leads to a new, very important conclusion: the energy of a particle at rest is equal to mc^2 . It is natural to call the quantity mc^2 the rest energy. Every particle possessing a rest mass m has also a rest energy mc^2 .

The energy of a moving particle can be connected with the relativistic mass by the relation

$$E = m(v)c^2, \quad (13.15)$$

analogous to (13.13).

Formulae (13.13) and (13.15), often called Einstein's formulae, show that every particle possessing a mass m has, at the same time, an energy E . The energy and mass are indissolubly connected with each other and are proportional to each other. This statement is often called the principle of equivalence of mass and energy.

Of course, the notions of equivalence and identity should not be confused. The energy and mass are different physical characteristics of particles, and the principle of equivalence established only their proportionality to each

other. The relation between the mass and energy is similar to the relation between the gravitational and inertial mass in classical mechanics: the two masses are indissolubly connected with each other and proportional to each other, but are at the same time different characteristics. *

At present Einstein's formula (13.15) as well as formulae (13.2) and (13.3) are confirmed by a vast amount of experimental material and their validity is beyond doubt.

If the particle is not acted upon by any external forces, then the energy conservation law

$$\frac{dE}{dt} = 0 \quad \text{or} \quad E = \text{const}, \quad (13.16)$$

and the momentum conservation law

$$\frac{dp}{dt} = 0 \quad (13.17)$$

hold. From Einstein's formula it follows that with the energy conservation law the law of conservation of the relativistic mass

$$m(v) = \text{const} \quad (13.18)$$

automatically holds.

As distinct from classical physics, where there are two independent conservation laws (i.e. the energy conservation law and the mass conservation law) there is in the theory of relativity only one conservation law, the law of conservation of energy or relativistic mass.

We shall come back to the physical interpretation of the energy conservation law in the theory of relativity in §15–17.

In conclusion we note that E is usually called the total energy. This should not lead to misunderstanding: E does not include the potential energy of the particle in an external field, if such a field acts on the particle. Sometimes one introduces the kinetic energy E_{kin} defining it as the energy of motion of the particle, i.e.

$$E_{\text{kin}} = E - mc^2 = mc^2 \left[\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} - 1 \right] = [m(v) - m]c^2. \quad (13.19)$$

* For more detail see: V.A.Fock, *The theory of space, time and gravitation* (Pergamon, London, 1964).

We introduce one more useful formula, connecting the energy and the three-dimensional momentum. From (12.12) and (13.1) it follows directly that

$$p = Ev/c^2. \quad (13.20)$$

By means of formula (13.13) one can find the expression for the classical radius of electron r_0 , introduced in Part I.

We have already discussed in §18 of Part I the fundamental difficulty of classical field theory associated with the problem of the self energy of the electron. For the energy of the electron in its own field one obtains the expression

$$U = e/r_0,$$

which diverges as $r_0 \rightarrow 0$.

If it is assumed that the overall mass of the electron is associated with the energy of the field produced by it, i.e. if

$$U \sim mc^2$$

for the electron at rest, then for r_0 one obtains the value

$$r_0 \sim e^2/mc^2,$$

which is the same as the inequality (29.6) of Part I. However, it should be borne in mind that the problem of the self energy is not solved at all by introducing the radius of the electron.

We have stressed that in the theory of relativity no body whatever, including the electron, can be considered as a perfectly rigid ball of given radius. Hence the quantity r_0 cannot be interpreted as the true "radius" of the electron. It is the minimum dimension of the space region in which use can still be made of the relations of classical field theory, the limit of applicability of its concepts. We note that, as has already been said in §29 of Part I, the true limit of applicability of classical field theory sets in at considerably larger distances.

§ 14. Lagrange's equations; the Lagrangian and Hamiltonian

As in classical mechanics, equations of motion can be written in generalized coordinates in the form of Lagrange's equations. For this we must first of all find the Lagrangian.

By definition the Lagrangian is a quantity whose derivatives with respect to the components of velocity represent the components of the momentum, while its derivatives with respect to coordinates represent the components of the force.

Hence we have

$$\frac{\partial L}{\partial v_i} = p_i = \frac{mv_i}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (14.1)$$

$$\frac{\partial L}{\partial x_i} = \mathcal{F}_i = - \frac{\partial U_{\text{field}}}{\partial x_i}, \quad (14.2)$$

where U_{field} is the potential energy of the external field, depending only on the coordinates of the particle. Eqs. (14.1) and (14.2) are satisfied by the Lagrangian:

$$L = -mc^2 \sqrt{1 - \frac{v^2}{c^2}} - U_{\text{field}} \quad (14.3)$$

We see that in relativistic mechanics the Lagrangian no longer represents the difference between the kinetic and potential energy.

By means of the Lagrangian one can write Lagrange's equations in generalized coordinates q_i :

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0, \quad (14.4)$$

or

$$\frac{d}{dt} p_i - \frac{\partial L}{\partial q_i} = 0,$$

where

$$p_i = \frac{\partial L}{\partial \dot{q}_i}$$

are generalized momenta.

Knowing the Lagrangian one can then find the Hamiltonian H . If the Lagrangian does not depend explicitly on time, then, by definition,

$$\begin{aligned}
 H = \sum q_i p_i - L &= \frac{mv^2}{\sqrt{1 - \frac{v^2}{c^2}}} + mc^2 \sqrt{1 - \frac{v^2}{c^2}} + U_{\text{field}} = \\
 &= \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} + U_{\text{field}} .
 \end{aligned} \tag{14.5}$$

The velocity can be expressed in terms of the momentum by means of the relations (12.2), from which it follows that

$$p_x^2 + p_y^2 + p_z^2 = p^2 = \frac{m^2 v^2}{1 - \frac{v^2}{c^2}}$$

A simple calculation gives

$$H = \sqrt{p^2 c^2 + m^2 c^4} + U_{\text{field}} . \tag{14.6}$$

Formula (14.6) shows that in relativistic mechanics the Hamiltonian is just the total energy expressed in terms of the momentum of the particle.

For large values of the momentum, when $p \gg mc$, the expression for the Hamiltonian of a free particle attains a simple form:

$$H \approx pc . \tag{14.7}$$

A motion with such a large momentum, for which the approximate formula (14.7) is valid, is called an ultrarelativistic motion. It is clear that for particles with zero rest mass formula (14.7) is valid. We shall often make use of it in considering processes in which light quanta (photons) take part.

§15. The mechanics of a system of particles in the theory of relativity

Up to now we have restricted ourselves to the consideration of one particle. The formulation of the mechanics of a system of particles in the theory of relativity is a much more complex problem. Nevertheless, in this case also a number of important general laws can be established.

A system of particles as a whole can be characterized by its energy E , momentum \mathbf{P} and rest mass M . If we are interested in the motion of the system as a whole, then, disregarding internal processes in the system and its spatial extension, the system can be considered as one material point. For the system as a whole one can write the equality

$$E = Mc^2, \quad (15.1)$$

where M is the rest mass of the entire system. Since it is always true that $M > 0$, the energy of a system of particles, like the energy of a single free particle, is an essentially positive quantity.

However, in general it is impossible to find expressions for the energy and momentum of the system in terms of the corresponding quantities for its individual particles, or to find general relations between the energy and momentum. The interaction existing between the particles can lead, for example, to the dependence of E on time (it should be recalled that E denotes the sum of the rest energy and kinetic energy, but does not include the energy of interaction). Hence the actual construction of the mechanics of a system of particles is restricted to a relatively few simple cases. These are:

- 1) a system of non-interacting particles,
- 2) a system of particles which are at large distances from one another and are moving with very large velocities,
- 3) systems of particles with a weak electromagnetic interaction.

The last case will be considered later, in § 25.

In a system of non-interacting particles the energy and momentum possess additive properties, so that

$$E = \sum_{i=1}^N E_i = \sum_{i=1}^N \frac{m_i c^2}{\sqrt{1 - \frac{v_i^2}{c^2}}}, \quad (15.2)$$

$$\mathbf{P} = \sum_{i=1}^N \mathbf{p}_i = \sum_{i=1}^N \frac{m_i \mathbf{v}_i}{\sqrt{1 - \frac{v_i^2}{c^2}}}, \quad (15.3)$$

where N is the number of particles in the system, and the index i refers to every one of the particles. The velocities of all particles are constant and, consequently, the total energy and total momentum of the system are also constant in time.

It is easily seen that the quantities E and \mathbf{P} form a 4-vector of energy and

momentum. Indeed, one can write

$$E = \sum_{i=1}^N \frac{m_i c^2}{\sqrt{1 - \frac{v_i^2}{c^2}}} = \text{const} , \quad (15.4)$$

$$\mathbf{P} = \sum_{i=1}^N \frac{m_i \mathbf{v}_i}{\sqrt{1 - \frac{v_i^2}{c^2}}} = \text{const} . \quad (15.5)$$

Hence the 4-vector

$$P_\alpha = \sum_{i=1}^N p_\alpha^{(i)} \quad (15.6)$$

can be introduced. Every term of the sum $p_\alpha^{(i)}$ is the 4-vector of the energy and momentum of an individual particle.

Before passing on to the discussion of the consequences of these propositions, we shall show that in the very important second case cited above the properties of a system of interacting particles can be reduced to those of non-interacting particles.

Consider a system of particles moving at large distances from one another. By the assumption that the velocities of the particles are very large, of the order of magnitude of the velocity of light, the energy

$$E_i = \frac{m_i c^2}{\sqrt{1 - \frac{v_i^2}{c^2}}}$$

is in general large in comparison with the interaction energy at large distances. Hence it can be assumed that, as for free particles, the total energy and momentum of the system are expressed by formulae (15.4) and (15.5).

When particles approach each other (this situation is often called a collision) the interaction between the particles can become appreciable and formulae (15.4) and (15.5) lose their applicability. However, when particles again diverge to large distances after collision the formulae (15.4) and (15.5) are again applicable. It is obvious, moreover, that the total energy and momentum of the particles which have diverged cannot differ from the energy and momentum before the interaction. Hence the conservation laws can be written in the form

$$\sum_{i=1}^N \frac{m_i c^2}{\sqrt{1 - \frac{v_i^2}{c^2}}} = \sum_{k=1}^{N^*} \frac{m_k c^2}{\sqrt{1 - \frac{v_k^2}{c^2}}}, \quad (15.7)$$

$$\sum_{i=1}^N \frac{m_i \mathbf{v}_i}{\sqrt{1 - \frac{v_i^2}{c^2}}} = \sum_{k=1}^{N^*} \frac{m_k \mathbf{v}_k}{\sqrt{1 - \frac{v_k^2}{c^2}}}. \quad (15.8)$$

Here the indices i and k refer to particles before and after the interaction. The asterisk in the sums on the right underlines the fact that the number of particles in the system before and after the interaction can be different (see § 17).

The quantities E and \mathbf{P} , referring to particles before and after the interaction, form a 4-vector of energy and momentum. According to (13.11), one can write the invariant for this 4-vector:

$$I = P_\alpha^2 = \sum_\alpha \left(\sum_i p_\alpha^{(i)} \right)^2 = \text{invar} = -M^2 c^2. \quad (15.9)$$

The constancy of the invariant I expresses the general energy-momentum conservation law.

We transform this invariant to a form convenient for practical use. We have

$$\begin{aligned} \sum_\alpha \left(\sum_i p_\alpha^{(i)} \right)^2 &= \sum_{i, \alpha} (p_\alpha^{(i)})^2 + 2 \sum_{k < i, \alpha} p_\alpha^{(i)} p_\alpha^{(k)} = \\ &= \sum_i (\mathbf{p}_i)^2 - \sum_i \frac{E_i^2}{c^2} + 2 \sum_{k < i} \mathbf{p}_i \mathbf{p}_k - 2 \sum_{k < i} \frac{E_i E_k}{c^2}. \end{aligned}$$

Here we have carried out the summation over the index α of the 4-vector and made use of formula (13.4) for $p_\tau^{(i)}$. Regrouping the terms in the sum, we find

$$\begin{aligned} I &= \left[\sum_i (\mathbf{p}_i)^2 + 2 \sum_{k < i} \mathbf{p}_i \mathbf{p}_k \right] - \frac{1}{c^2} \left[\sum_i E_i^2 + 2 \sum_{k < i} E_i E_k \right] = \\ &= \left(\sum_i \mathbf{p}_i \right)^2 - \frac{1}{c^2} \left(\sum_i E_i \right)^2 = \mathbf{P}^2 - \frac{E^2}{c^2} = -M^2 c^2. \quad (15.10) \end{aligned}$$

Thus, we can write

$$P^2 - \frac{E^2}{c^2} = (P^*)^2 - \frac{(E^*)^2}{c^2}, \quad (15.11)$$

where the values of the quantities after the interaction are denoted by an asterisk. The components of the 4-vector P_α transform, in the transition from one inertial reference frame to another, according to the general formulae (13.6)–(13.9).

It often turns out to be convenient to make use of a reference frame in which the total momentum is equal to zero. Such a reference frame, as in classical mechanics, is called the centre of mass system $K^{(c.m.)}$.

Let a reference frame K be given in which a system of particles has a momentum P and an energy E . Let us find the velocity of motion $V^{(c.m.)}$ of the centre of mass of the system with respect to the reference frame K . The velocity $V^{(c.m.)}$ characterizes the motion of the system as a whole. For simplicity we assume that the velocity of the centre of mass of the system is directed along the x -axis. Then according to the inverse of (13.6)–(13.9) we have

$$P_x^{(c.m.)} = 0 = \frac{P_x - V^{(c.m.)} \frac{E}{c^2}}{\sqrt{1 - \frac{(V^{(c.m.)})^2}{c^2}}}, \quad (15.12)$$

$$P_y^{(c.m.)} = 0 = P_y, \quad P_z^{(c.m.)} = 0 = P_z, \quad (15.13)$$

$$E^{(c.m.)} = \frac{E - V^{(c.m.)} P_x}{\sqrt{1 - \frac{(V^{(c.m.)})^2}{c^2}}}. \quad (15.14)$$

Quantities referred to the centre of mass system are denoted by the superscript $c.m.$.

From (15.12) we obtain

$$V^{(c.m.)} = \frac{P_x c^2}{E}. \quad (15.15)$$

In the general case of an arbitrary orientation of the velocity vector of the centre of mass system, instead of (15.15) one obtains

$$\mathbf{V}^{(c.m)} = \frac{\mathbf{P}c^2}{E}. \quad (15.16)$$

In contrast to classical mechanics, the velocity of the centre of mass in relativistic mechanics cannot be represented as the derivative of the coordinate of the centre of mass with respect to time:

$$\mathbf{V}^{(c.m)} \neq \frac{d}{dt} \mathbf{R}^{(c.m)}.$$

The quantity $\mathbf{P}c^2/E$ cannot, in general, be expressed in the form of the derivative of any quantity with respect to time. Hence it is impossible in relativistic mechanics to introduce the notion of the coordinate of the centre of mass for an arbitrary system of interacting particles. In §25 it will be shown that in the case of a system of weakly interacting particles use can be made, in a certain approximation, of the concept of the centre of mass.

We shall also discuss some properties of the rest mass of a system of particles.

Consider a centre of mass system K' . We write the invariant of the 4-vector of energy and momentum (15.10) for this system in the form

$$(\mathbf{P}^{(c.m)})^2 c^2 - (E^{(c.m)})^2 = -M^2 c^4. \quad (15.17)$$

Since $\mathbf{P}^{(c.m)} = 0$, formula (15.17) in the system K' assumes the form

$$E^{(c.m)} = Mc^2, \quad (15.18)$$

where M is the rest mass.

On the other hand, according to the general formula (15.2),

$$E^{(c.m)} = \sum_{i=1}^N \frac{m_i c^2}{\sqrt{1 - \frac{v_i^2}{c^2}}},$$

where v_i is the velocity of the i th particle referred to the centre of mass system. Then for the rest mass we obtain

$$M = \sum_{i=1}^N \frac{m_i}{\sqrt{1 - \frac{v_i^2}{c^2}}} \quad (15.19)$$

We see that the rest mass of a system of particles is not equal to the sum of the rest masses of individual particles $\sum m_i$, but depends on the velocities of their motion with respect to the centre of mass system.

Consider, for example, the case where a system of particles represents an ideal gas. The ideal gas obviously satisfies the requirements regarding the character of the interaction between the particles about which we have spoken earlier. Then, according to (15.17), the rest mass M of the gas depends on the internal motion of the gas particles or, what is the same, on the temperature of the gas.

The difference between the rest mass M of the system and the sum $\sum m_i$ of the rest masses of the particles forming the system is of great significance for processes occurring in nature. We shall discuss this in the next paragraph.

In the general case of a system of particles with an arbitrary interaction formula (15.2) is not an expression for the total energy of the system; it gives only the sum of the rest masses and kinetic energies of the particles.

It is necessary to include the energy of interaction between the particles in the total energy. It turns out, however, that in relativistic mechanics the idea of a potential energy of interaction of a system of particles does not exist. The potential energy of interaction of particles must depend only on their position. If the position of any particle is changed, then the potential energy of the system of particles and the forces acting on individual particles must change instantly. In other words, the concept of a potential energy of interaction of a particle is associated with the concept of action at a distance and cannot be admitted in the theory of relativity. In general, it appears to be impossible to write an expression for the energy of a system of interacting particles. The same holds also for the momentum of the system, which in the theory of relativity is not a time-independent quantity.

Besides systems interacting via collisions, an approximate expression can be found for the interaction of charged particles in the special theory of relativity. This will be done in §25. The gravitational interaction of bodies is considered in the general theory of relativity the subject of which is beyond the scope of our book.

§16. The energy-momentum conservation law in nuclear physics

The energy-momentum conservation law and the mass-energy relation have not only found experimental confirmation, but have become basic propositions in contemporary nuclear physics. As far back as the first work of Ein-

stein it was pointed out that the mass-energy relation could be checked experimentally in investigating the phenomena of radioactivity. Indeed, a characteristic feature of radioactive decay as, after all, of all other nuclear processes, is a large change in the energy of the system and the high energies of the nuclear particles produced. Of the whole variety of nuclear processes in which relativistic effects play an essential role, we have to confine ourselves within the framework of this book to only the most essential or typical ones.

The purpose of the examples presented below is to show that the relations of relativistic mechanics are a necessary basis for a study of processes taking place with atomic nuclei and elementary particles.

1. *Reactions involving the decay of particles.* A number of basic processes occurring with atomic nuclei and elementary particles arise in fusion reactions and the decay of particles. The energy-momentum conservation law imposes a fundamental restriction on the possible reactions. Consider the decay of one particle or body into two others. We shall assume that the decay takes place spontaneously, i.e. as a result of internal changes in the system, without the action of external forces on the system.

Let the disintegrating particle have a mass M . In the centre of mass system the momentum before the decay is zero. The invariant of the 4-vector of energy and momentum is equal to

$$I = \mathbf{p}^2 c^2 - E^2 = -M^2 c^4$$

After the decay two particles arise with masses m_1, m_2 , momenta $\mathbf{p}_1, \mathbf{p}_2$ and energies E_1, E_2 .

In the centre of mass system the total momentum after the decay is zero:

$$\mathbf{p}_1 + \mathbf{p}_2 = 0.$$

The invariant I can be written in the form

$$I = (\mathbf{p}_1 + \mathbf{p}_2)^2 c^2 - (E_1 + E_2)^2 = -M^2 c^4,$$

or

$$Mc^2 = E_1 + E_2.$$

Writing the energies E_1 and E_2 in the form

$$E_1 = m_1 c^2 + E_{\text{kin}}^{(1)},$$

$$E_2 = m_2 c^2 + E_{\text{kin}}^{(2)},$$

we have

$$Mc^2 = (m_1 + m_2)c^2 + E_{\text{kin}}^{(1)} + E_{\text{kin}}^{(2)}. \quad (16.1)$$

Since the kinetic energies of the particles produced after the decay are each greater than zero, it follows from (16.1) that the spontaneous decay of a body is possible only if the inequality

$$M > m_1 + m_2 \quad (16.2)$$

is fulfilled, i.e. only if the mass of the body is larger than the sum of the rest masses of the decay products. Conversely, in the cases when the mass of the body is smaller than the sum of the masses of the decay products, a spontaneous decay of the body is impossible. For the decay to occur in this case it is necessary to supply energy from outside.

Making use of the equalities $E_1 = \sqrt{p_1^2 c^2 + m_1^2 c^4}$ and $E_2 = \sqrt{p_2^2 c^2 + m_2^2 c^4}$ and taking into account that $p_1^2 = p_2^2$, one can easily find the energy of the particles produced in the decay.

Namely, we have

$$E_2^2 = p_2^2 c^2 + m_2^2 c^4 = E_1^2 + m_2^2 c^4 - m_1^2 c^4$$

On the other hand,

$$E_1^2 = (Mc^2 - E_2)^2,$$

hence

$$E_1 = \frac{(M^2 - m_2^2 + m_1^2)c^2}{2M}, \quad E_2 = \frac{(M^2 - m_1^2 + m_2^2)c^2}{2M}$$

2. Stability of atomic nuclei. The general results obtained above allow one to clarify the very important problem of stability of atomic nuclei.

Consider an atomic nucleus consisting of Z protons and $A - Z$ neutrons (where Z is the atomic number, and A is the mass number) and having a mass M . The protons and neutrons in the nucleus possess considerable kinetic energies. However, strong attractive forces — nuclear forces — acting between them ensure the stability of the system as a whole. The rest energy of the nucleus Mc^2 is made up of the rest energy of all particles constituting it, $\sum m_i c^2$, and the energy of internal motion and interaction of the particles.

In order for a nucleus to be stable and that the motion of nuclear particles should not lead to its spontaneous decay, it is obviously necessary that the inequality

$$Mc^2 < \sum m_i c^2 \quad (16.3)$$

be fulfilled. The quantity

$$\Delta mc^2 = \sum m_i c^2 - Mc^2, \quad (16.4)$$

called the nuclear binding energy, is a measure of the stability of the nucleus. If, in particular, Δmc^2 is negative, the nucleus is unstable and decays spontaneously.

In addition to the binding energy, the quantity

$$\Delta m = \sum m_i - M, \quad (16.5)$$

called the mass defect, serves as a measure of the stability of the nucleus. In order for a nucleus to be stable it is necessary that the mass defect be positive. If the mass defect of a nucleus is positive, then according to (16.1) the nucleus is stable with respect to decay into its constituent particles — protons and neutrons. However, this does not mean that the nucleus is absolutely stable and that it can exist for an indefinitely long time.

Assume that as a result of a decay of a nucleus, for example according to the scheme

$$M \rightarrow M_1 + M_2,$$

nuclei with masses M_1 and M_2 are produced. Such a decay is in principle possible if the atomic numbers Z_1 and Z_2 and mass numbers A_1 and A_2 of the product nuclei satisfy the equalities

$$Z_1 + Z_2 = Z, \quad A_1 + A_2 = A.$$

Let Δm_1 and Δm_2 be the mass defects of the product nuclei. If the mass defect Δm of the initial nucleus is smaller than the sum of the mass defects of the product nuclei, i.e.

$$\Delta m < (\Delta m_1 + \Delta m_2),$$

then the system arising after the decay possesses a higher stability than the

initial system. Hence when $\Delta m < (\Delta m_1 + \Delta m_2)$ the nucleus, which is stable with respect to the decay into individual elementary particles, is not stable with respect to the decay into two parts. As a result of transformations of the nuclear particles, a decay configuration will arise in the nucleus after the lapse of a certain time interval. The initial nucleus then decays into two nuclei with masses M_1 and M_2 .

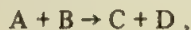
As an example the nucleus ${}^8_4\text{Be}$ can be considered. This nucleus has a mass $M = 8.005\,31$ amu, which is smaller than the sum of the masses of four protons and four neutrons: $\sum m_i = 4 \times 1.007\,825$ amu + $4 \times 1.008\,665$ amu = $8.065\,960$ amu. Hence the nucleus ${}^8_4\text{Be}$ is stable with respect to decay into individual protons and neutrons. However, the mass of ${}^8_4\text{Be}$ is larger than the mass of two nuclei of ${}^4_2\text{He}$: $2M_{\text{He}} = 2 \times 4.002\,603$ amu = $8.005\,206$ amu. Hence the nucleus ${}^8_4\text{Be}$ is unstable and must decay spontaneously into two α -particles, which indeed occurs.

On the contrary, the mass defect of the nucleus ${}^9_4\text{Be}$ is not only positive but also exceeds the sum of the mass defects of all nuclei into which it could decay. Hence the nucleus ${}^9_4\text{Be}$ is absolutely stable.

Knowing the masses of all isotopes one can easily determine their stability from the values of the mass defects.

For details we refer the readers to more specialized texts.

3. Energy yield of nuclear reactions. The application of the energy conservation law to a nuclear reaction of the type

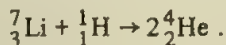


where A and B are the initial nuclei, and C and D are the reaction products, allows one to find the energy yield of the reaction

$$Q = [(M_A + M_B) - (M_C + M_D)]c^2, \quad (16.6)$$

if the masses of all the nuclei are known.

As an important example from which the relations of the theory of relativity can be checked in a particularly obvious way, we consider the reaction



The masses of all the nuclei figuring in the reaction have been measured with a high degree of accuracy.

$$M({}^7_3\text{Li}) = 7.016\,004 \text{ amu},$$

$$M({}^1_1\text{H}) = 1.007\,825 \text{ amu}.$$

The total initial mass is 8.023 829 amu. The total final mass is $2M({}_2^4\text{He}) = 2 \times 4.002\,603 = 8.005\,206$ amu. As a result of the reaction the rest mass of the particles is reduced by

$$\Delta m = 0.018\,623 \text{ amu.}$$

The corresponding energy, representing the kinetic energy of the two α -particles, must be equal to (1 amu = 931.48 MeV)

$$E = \Delta mc^2 = 17.3 \text{ MeV,}$$

which agrees to a high degree of accuracy with the measured values of the energy.

In this example it is seen that the rest mass of the particles is not conserved.

In the course of the reaction a rest mass equal to Δm vanishes. However the energy conservation law

$$E_{\text{Li}} + E_{\text{H}} = 2E_{\text{He}}$$

must hold and also the relativistic mass conservation law

$$m_{\text{Li}}(v) + m_{\text{H}}(v) = 2m_{\text{He}}(v).$$

The above reaction is only one example of nuclear reactions in which the rest mass conservation law is not fulfilled.

4. *Decay of elementary particles.* A number of elementary particles turn out to be unstable with respect to decay. It is impossible to discuss here all the known cases of decay, and we can dwell only on some of them, illustrating in an obvious way the importance of the application of conservation laws to the analysis of decay.

As an example we shall consider the decay of charged and neutral mesons. As is well known, the existence of three kinds of π -mesons (pions) has been established: π^+ with a positive charge, π^- with a negative charge and π^0 , the neutral meson. The charged mesons have masses of $273m_e$, and the neutral meson has a mass of $264m_e$. The charge of the charged mesons is equal in its absolute value to that of an electron. In addition to π -mesons, two kinds of μ -mesons (muons) have been discovered: the positive μ^+ and the negative μ^- with the same absolute value of the charge and a mass equal to $207m_e$.

It turns out that both π -mesons and μ -mesons are unstable. The lifetime of

charged π^\pm -mesons is 2.6×10^{-8} sec, and that of the neutral π^0 -meson is about $10^{-15} - 10^{-16}$ sec. As mentioned earlier, μ^\pm -mesons have a lifetime of 2×10^{-6} sec.

It is clear that the establishment of the decay scheme of all these particles is of fundamental importance for contemporary physics.

We begin with the decay of charged π -mesons. A charged π^\pm -meson decays into a μ^\pm -meson and a neutrino ν :

$$\pi^\pm \rightarrow \mu^\pm + \nu.$$

The neutrino is a neutral particle with a very small rest mass, which, in contrast to photons, does not cause any ionization of atoms in its passage through matter. The study of the above reaction allows one to estimate most accurately the rest mass of the neutrino. *

That is to say, knowing the rest masses of π -mesons and μ -mesons one can find the kinetic energy of the μ -meson E_{kin}^μ as a function of the rest mass of the neutrino.

In the centre of mass system the momentum of the π -meson before decay is equal to zero. After decay the total momentum remains equal to zero, so that the momenta of the μ -meson and neutrino are of the same magnitude $p_\mu = p_\nu = p$ and of opposite direction, $(\mathbf{p}_\mu + \mathbf{p}_\nu) = 0$. The law of conservation of the energy-momentum 4-vector reduces to the energy conservation law

$$m_\pi c^2 = E^\mu + E^\nu,$$

or

$$m_\pi c^2 = (E_{\text{kin}}^\mu + m_\mu c^2) + \sqrt{p_\nu^2 c^2 + m_\nu^2 c^4},$$

or, since in the centre of mass system $p_\nu = p_\mu$, we have

$$m_\pi c^2 = (E_{\text{kin}}^\mu + m_\mu c^2) + \sqrt{p_\mu^2 c^2 + m_\nu^2 c^4} \quad (16.7)$$

We make use of the identity

$$E_{\text{kin}} + mc^2 = \sqrt{p^2 c^2 + m^2 c^4},$$

or

$$E_{\text{kin}}^2 + 2mc^2 E_{\text{kin}} = p^2 c^2, \quad (16.8)$$

* For the properties of the neutrino see § 123 of Part V.

and rewrite (16.7) in the form

$$m_{\pi}c^2 - (E_{\text{kin}}^{\mu} + m_{\mu}c^2) = \sqrt{(E_{\text{kin}}^{\mu})^2 + 2m_{\mu}c^2E_{\text{kin}}^{\mu} + m_{\nu}^2c^4}$$

Squaring the above formula we find for the energy of the μ -meson (referred to the centre of mass system) produced in the decay of a π -meson

$$E_{\text{kin}}^{\mu} = \frac{(m_{\pi}c^2 - m_{\mu}c^2)^2 - m_{\nu}^2c^4}{2m_{\pi}c^2} \quad (16.9)$$

The kinetic energy, E_{kin}^{μ} , of the μ -meson may be measured from the ionization caused by it. It turns out always to have a unique value. From the measured value of E_{kin}^{μ} it follows that, within the limits of experimental accuracy, $m_{\nu} = 0$. Moreover, the fixed value of E_{kin}^{μ} points out the correctness of the assumed decay scheme. If two or more neutral particles appeared in the decay, then E_{kin}^{μ} would not have a well defined value but would depend on the distribution of the kinetic energy between the neutral particles.

This is just the case realized in the decays of μ -mesons, which proceed according to the scheme

$$\mu^{\pm} \rightarrow e^{\pm} + 2\nu.$$

Here e^{-} stands for an electron, and e^{+} stands for a positron.

The theory of positrons will be discussed in detail in Part V. Here we point out only that the positron has a positive charge of the same magnitude as the electron charge and a mass equal to that of the electron. The electron, π^{-} -meson and μ^{-} -meson on the one hand, and the positron, π^{+} -meson and μ^{+} -meson on the other hand form groups of antiparticles. Antiparticles are produced in pairs which can be annihilated, producing γ -quanta. *

Experiment shows that in the decay of μ -mesons the kinetic energy of the electrons or positrons produced has no well defined value but varies from one decay to another.

Thus, it is clear that the energy carried away by neutral particles escaping direct observation can be distributed between them in different ways. This

* For the properties of elementary particles see: Y.V. Novozhilov, *Elementary particles* (Gordon and Breach, New York, 1961) (a popular and very good exposition), and E.V. Shpol'skii, *Atomnaya fizika* (Atomic particles), Vol. II (Gostekhizdat, Moscow, 1951).

would be impossible for a decay according to the scheme which is valid for π -mesons, i.e. with the production of one neutrino.

The kinetic energy of the charged particle is on the average equal to $\frac{1}{3}$ of the total energy of the μ -meson. The remaining $\frac{2}{3}$ are on the average equally distributed between the two neutrinos.

Consider, finally, the decay of the π^0 -meson, which proceeds according to the scheme

$$\pi^0 \rightarrow \gamma + \gamma,$$

where γ stands for a photon. γ -quanta are detected by the ionization they produce. A certain relation is observed between the angle of divergence of the photons and their energy. This relation can be established from the conservation laws.

Before the decay the invariant of the 4-vector of energy and momentum in a reference frame moving together with the π^0 -meson reduces to

$$I = -m_\pi^2 c^4,$$

since in this reference frame the total momentum is equal to zero.

After the decay in the laboratory system I is equal to

$$I = (\mathbf{p}_1 + \mathbf{p}_2)^2 c^2 - (E_1 + E_2)^2,$$

where $\mathbf{p}_1, \mathbf{p}_2, E_1, E_2$ are the momenta and energies of the two photons. Hence

$$-m_\pi^2 c^4 = (\mathbf{p}_1 + \mathbf{p}_2)^2 c^2 - (E_1 + E_2)^2,$$

or, since the rest mass of photons is equal to zero,

$$\begin{aligned} m_\pi^2 c^4 &= (p_1 + p_2)^2 c^2 - (\mathbf{p}_1 + \mathbf{p}_2)^2 c^2 = \\ &= 2p_1 p_2 c^2 (1 - \cos \varphi) = 4p_1 p_2 c^2 \sin^2 \frac{1}{2} \varphi, \end{aligned}$$

where φ is the angle between the directions of flight of the photons.

Consequently,

$$\sin \frac{1}{2} \varphi = \frac{m_\pi c^2}{2\sqrt{E_1 E_2}}. \quad (16.10)$$

This dependence of the angle of divergence on the energy of the photons is in good agreement with that observed experimentally.

5. *Electron-positron pair production by γ -quanta and positron-electron annihilation.* The experimental discovery of the theoretically predicted phenomena of electron-positron pair production and positron-electron annihilation with the production of γ -quanta appeared to be a basic confirmation of the validity of relativistic quantum mechanics, which the reader will meet in Part V of the book. At the same time, these phenomena serve as good illustrations of the relations of the theory of relativity.

Consider first of all the question of the possibility of electron-positron pair production by a γ -quantum in vacuum. Let a γ -quantum with energy $E = pc$ produce an electron and a positron. It is easily seen that such a process is incompatible with conservation laws. Indeed, if the photon produces an electron-positron pair with the minimum possible energy — i.e. the rest energy — and a momentum equal to zero, then it has an energy $E = 2mc^2$ and a momentum $p = E/c > 0$, where m is the electron mass. If the momentum of the pair differs from zero, then it is always possible to transform to the centre of mass system where it is equal to zero, and our reasoning is still correct.

Thus, the momentum differs from zero before the pair production, and is equal to zero after the pair production. This would clearly contradict the momentum conservation law, and therefore such a process is impossible. Pair production can take place only in the presence of a third body; usually an atomic nucleus, which itself takes the excess momentum. Since the nucleus also acquires a certain part of the energy, the energy threshold of pair production by a γ -quantum lies above $2mc^2$. The threshold energy of the γ -quantum is determined by the requirement that all the particles — the electron, positron and nucleus — should have a momentum equal to zero in the centre of mass system.

To find this threshold we shall find the invariant of the energy-momentum 4-vector.

Before the reaction there was a γ -quantum with a threshold kinetic energy E_{thresh} and a momentum $p = E_{\text{thresh}}/c$ and a nucleus at rest with a mass M . After the reaction there was an electron-positron pair and the nucleus which acquired a part of the momentum and energy.

The value of the invariant of the energy-momentum 4-vector before the reaction was

$$I = \left(\frac{E_{\text{thresh}}}{c} \right)^2 c^2 - (E_{\text{thresh}} + Mc^2)^2$$

The value of the invariant after the reaction can be taken in any reference frame, including the centre of mass system. The merit of the latter is the fact that in it the total momentum of all the particles is equal to zero and the value of the invariant reduces to

$$I = -(Mc^2 + 2mc^2)^2.$$

Thus, we have

$$(Mc^2 + E_{\text{thresh}})^2 - E_{\text{thresh}}^2 = (Mc^2 + 2mc^2)^2.$$

Hence for the threshold energy of the γ -quantum we find

$$E_{\text{thresh}} = 2mc^2 \left(1 + \frac{m}{M} \right) \quad (16.11)$$

The threshold energy for other processes producing electron-positron pairs, for example the collision between two electrons, can be found in a completely analogous way.

The process which is the reverse of the process of pair production by γ -quanta is called pair annihilation. In annihilation an electron and a positron fuse to produce γ -quanta. The essence of this process from the point of view of contemporary quantum theory will be discussed briefly in Part V. The process of annihilation usually takes place at small values of kinetic energy of the positron. Hence the difference between the energy of the initial state and the energy of the final state amounts to

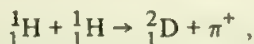
$$\Delta E = mc^2 - (-mc^2) = 2mc^2 = 1.02 \text{ MeV}$$

The simultaneous conservation of energy and momentum requires that this energy should be emitted in the form of (at least) two γ -quanta emerging in opposite directions and having an energy of 0.51 MeV each.

It has been established in a number of experiments that the annihilation of positrons is indeed accompanied by such radiation. The phenomenon of the annihilation of positrons serves as one of the most effective confirmations of the mass-energy relation. It should be noted that the process of annihilation is sometimes incorrectly interpreted as "the transformation of matter into energy" or as "the disappearance of particles". It is clear that the γ -quanta arising in the annihilation of a positron are as material as the electron and positron.

In § 13 we have already discussed the properties of the rest mass and conservation laws in the theory of relativity and have seen that the transformation of particles with a rest mass different from zero into particles having no rest mass can in no way be considered as the disappearance of particles or the transformation of mass into energy.

6. As the last example of the use of conservation laws in nuclear physics we shall consider the determination of the energy threshold of the reaction in which a π -meson is produced in a proton-proton collision:



where π^+ is a positive π -meson with mass m_π .

We write the value of the invariant $E^2 - p^2c^2$ before and after the collision. Before the collision,

$$(E_{\text{thresh}} + 2Mc^2)^2 - p^2c^2 = (E_{\text{thresh}} + 2Mc^2)^2 - \\ - E_{\text{thresh}}(E_{\text{thresh}} + 2Mc^2).$$

Here we have assumed that one of the protons was at rest before the collision and made use of formula (16.8). If the incident proton has the threshold energy, then the particles produced in the reaction, the deuteron and π^+ -meson, have the minimum possible energy, which corresponds to the rest energy in the centre of mass system. In the centre of mass system one can write

$$E^2 - p^2c^2 = (2Mc^2 + m_\pi c^2)^2.$$

From the conservation of the invariant $(E^2 - p^2c^2)$ we have

$$(E_{\text{thresh}} + 2Mc^2)^2 - E_{\text{thresh}}(E_{\text{thresh}} + 2Mc^2) = (2Mc^2 + m_\pi c^2)^2$$

Hence the threshold value of the kinetic energy is

$$E_{\text{thresh}} = \frac{[(2M + m_\pi)^2 - 4M^2]c^2}{2M} = \frac{(4m_\pi M + m_\pi^2)c^2}{2M} = \\ = m_\pi c^2 \left(2 + \frac{m_\pi}{2M} \right) = 292 \text{ MeV}. \quad (16.12)$$

This value of E_{thresh} is in good agreement with the measured value.

The examples quoted are of a purely illustrative character, and have been used intentionally to introduce the diverse problems of nuclear physics. They show, however, that in all nuclear processes in which rather considerable changes in the energy of the system are to be taken into account the laws of the theory of relativity and, in particular, the mass-energy relation play a fundamental role.

§17. The theory of collisions between relativistic particles. Compton effect

The theory of collisions between relativistic particles is of great importance for nuclear physics. In the absence of nuclear reactions between the colliding particles the interaction between them can, with a sufficient degree of accuracy, be considered as an elastic collision (i.e. a collision without a change in the internal state of the nuclear particles). This refers, in particular, to collisions between elementary particles, for example, mesons, protons or photons with electrons.

We shall consider first of all the elastic collisions of particles with a rest mass differing from zero. We assume that a fast particle with mass μ and momentum \mathbf{p} collides with another particle having a mass m . We consider the second particle to be at rest and free. Such an approximation is legitimate for a sufficiently large velocity of the incident particle. After collision the particle which was initially at rest will move with a momentum \mathbf{p}_1 directed at an angle φ to the momentum of the incident particle. The latter will be deflected from the initial direction of flight by an angle θ and will have a momentum \mathbf{p}_2 (fig. II.4).

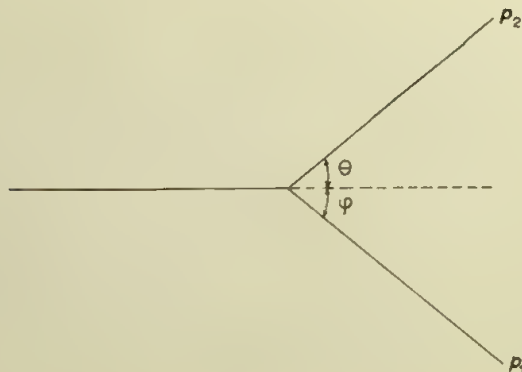


Fig. II.4

We can write the energy and momentum conservation laws in the form of (15.7) and (15.8):

$$\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2, \quad (17.1)$$

$$mc^2 + E = E_1 + E_2, \quad (17.2)$$

where E and E_2 represent the energy of the incident particle before and after collision respectively, and E_1 is the energy after collision of the particle which was initially at rest. These relations are sufficient to find all of the quantities characterizing the process of collision.

Suppose, for example, that we have to find the energy acquired by the particle which was initially at rest as a function of the angle φ . It is simpler to find, not the total energy E_1 of the particle, but its kinetic energy $E_1^{(\text{kin})} = E_1 - mc^2$. From (17.2) we have, obviously,

$$\sqrt{p^2 c^2 + \mu^2 c^4} = \sqrt{p_2^2 c^2 + \mu^2 c^4} + E_1^{(\text{kin})} \quad (17.3)$$

Since on the basis of (17.1)

$$p_2^2 = p^2 + p_1^2 - 2pp_1 \cos \varphi, \quad (17.4)$$

we can eliminate p_2 from (17.3) and (17.4). Moreover, p_1^2 can be expressed in terms of E_1 by means of (16.8). Rearranging and squaring (17.3) and substituting the value for p_2^2 from (17.4), we obtain after some manipulation

$$pp_1 c^2 \cos \varphi = E_1^{(\text{kin})} \{ \sqrt{p^2 c^2 + \mu^2 c^4} + mc^2 \}.$$

Squaring the above relation and expressing p_1^2 again in terms of $E_1^{(\text{kin})}$, we finally obtain

$$E_1^{(\text{kin})} = \frac{2mp^2 c^4 \cos^2 \varphi}{\{ \sqrt{p^2 c^2 + \mu^2 c^4} + mc^2 \}^2 - p^2 c^2 \cos^2 \varphi}, \quad (17.5)$$

i.e. the dependence on p and φ of the energy transferred to the motionless particle, as well as the dependence on the masses m and μ of the particles.

Formula (17.5) shows that the energy transferred has its largest value for $\varphi = 0$, i.e. for the motion of the initially stationary particle in the direction of flight of the incident particle (i.e. head-on collision). Then,

$$(E_1)_{\max}^{(\text{kin})} = 2mc^2 \frac{p^2 c^2}{\{\sqrt{p^2 c^2 + \mu^2 c^4} + mc^2\}^2 - p^2 c^2}. \quad (17.6)$$

Consider some particular cases of formula (17.6). Let, for example, the incident particle be a proton or a meson, and let the particle at rest be an electron. Then $\mu \gg m$. In addition, we assume the incident particle to be very fast, so that $pc \gg \mu c^2$. Then from (17.6) we find

$$\begin{aligned} (E_1)_{\max}^{(\text{kin})} &= 2mc^2 \frac{p^2 c^2}{\mu^2 c^4 + 2mc^2 \sqrt{p^2 c^2 + \mu^2 c^4} + m^2 c^4} \\ &\approx 2mc^2 \frac{p^2 c^2}{2mc^2 pc + \mu^2 c^4}. \end{aligned} \quad (17.6')$$

If the momentum of the incident particle is so large that the inequality

$$pc \gg \frac{\mu^2 c^4}{mc^2} = \frac{\mu}{m} \mu c^2$$

is fulfilled, then the maximum transferred energy amounts to

$$(E_1)_{\max}^{(\text{kin})} \approx pc \approx E,$$

i.e. to the energy of the incident particle.

Now consider the case when the incident particle is a light particle, for example, an electron, and the particle at rest is a heavy particle (i.e. $\mu \ll m$). If, moreover, the first particle possesses a momentum $pc \gg \mu c^2$, then from (17.6') we obtain for the energy transferred

$$(E_1)_{\max}^{(\text{kin})} \approx 2mc^2 \frac{p^2 c^2}{2mc^2 pc + m^2 c^4}.$$

If the inequality $pc \gg mc^2$ is satisfied, then

$$(E_1)_{\max}^{(\text{kin})} \approx pc \approx E. \quad (17.7)$$

Thus, we see that for very large momenta the laws for elastic collisions in relativistic mechanics differ in a fundamental way from the analogous laws in classical mechanics. For sufficiently large momenta a total energy transfer

from a heavy particle to a light one is possible, and also from a light particle to a heavy one. It should be noted that in this case in classical mechanics the elastic collision is accompanied by only a negligible energy transfer. *

Other limiting cases can be easily obtained from the general expression (17.5). In particular, it is easy to show that for small momenta $pc \ll \mu c^2$ and $pc \ll mc^2$ the energy transferred is given by a formula which is the same as the corresponding expression for the non-relativistic theory of collisions.

An important case of elastic collision is the collision of a photon with an electron. This phenomenon, called the Compton effect, was first studied carefully in connection with the elucidation of the quantum nature of light. The Compton effect plays a role in a number of practical problems of contemporary nuclear physics. The photon has a rest mass μ equal to zero and $E = pc$, $E_2 = p_2c$. Hence formulae (17.2) and (17.1) assume the forms

$$E = E_2 + E_1^{(\text{kin})}, \quad (17.8)$$

$$p_1^2 = \left(\frac{E}{c}\right)^2 + \left(\frac{E_2}{c}\right)^2 - \frac{2EE_2}{c^2} \cos \theta, \quad (17.9)$$

where $E_1^{(\text{kin})} = \sqrt{p_1^2 c^2 + m_e^2 c^4} - m_e c^2$, m_e is the mass of the electron, and θ is the angle between the direction of flight of the photon before and after collision (scattering angle).

Expressing $p_1^2 c^2$ in eq. (17.9) in terms of the kinetic energy of the electron by means of (16.8), we write (17.9) in the form

$$(E_1^{(\text{kin})})^2 + 2m_e c^2 E_1^{(\text{kin})} = E^2 + E_2^2 - 2EE_2 \cos \theta. \quad (17.10)$$

First of all we find the energy of the photon undergoing the collision. Eliminating $E_1^{(\text{kin})}$ from (17.8) and (17.10), we easily find

$$E_2 = \frac{m_e c^2 E}{m_e c^2 + E(1 - \cos \theta)}. \quad (17.11)$$

This formula relates the energy of the scattered photon to that of the incident one and to the scattering angle θ .

Transforming from energies to wavelengths according to the well-known quantum formula $E = h\nu = hc/\lambda$, one can obtain the following value for the decrease in the wavelength $\Delta\lambda$ in Compton scattering:

* See § 43 of Part I.

$$\Delta\lambda = hc \left(\frac{1}{E_2} - \frac{1}{E} \right) = \frac{h}{m_e c} (1 - \cos \theta) = \Lambda (1 - \cos \theta). \quad (17.12)$$

The quantity $\Lambda = h/m_e c = 0.0242 \text{ \AA}$ is called the Compton wavelength.

Formula (17.12) shows that the change in the wavelength is independent of the wavelength of the incident radiation. The maximum change in the wavelength is equal to 2Λ .

Knowing the energy of the scattered photon, one can easily find the energy transferred to the electron. It turns out to be

$$E_1^{(\text{kin})} = E - E_2 = \frac{\Lambda(1 - \cos \theta)}{\lambda + \Lambda(1 - \cos \theta)} E, \quad (17.13)$$

i.e. relatively not large for $\lambda \gg \Lambda$. On the contrary, for $\lambda \approx \Lambda$ the energy transferred to the electron turns out to be considerable, of the order of magnitude of E .

The importance of this result lies particularly in the fact that it is of a general character. We have more than once pointed out in Part I that classical electrodynamics contains in itself the limits of its applicability. Namely, classical electrodynamics become inapplicable in the region of small distances, of the order of magnitude of the classical radius of the electron $r_0 = 2.5 \times 10^{-13} \text{ cm}$. We have mentioned, however (see §29 of Part I), that in fact the limit of applicability of the classical theory lies much higher, at distances of the order of $2 \times 10^{-10} \text{ cm}$, which corresponds to the order of magnitude of the Compton wavelength.

The unsuitability of classical electrodynamics for phenomena taking place in a region of space with a linear dimension of the order of Λ is obviously associated with the fact that in this region quantum effects begin to make themselves felt. The scattering of light serves as a concrete example of this.

For $\lambda \gg \Lambda$ the change in the wavelength of the scattered light and the energy transferred to the free electron are relatively small. Hence the scattering of light is sufficiently well described by the classical theory. The scattering is coherent. The change in the wavelength $\Delta\lambda \approx \Lambda$ is very small in comparison with the wavelength itself; no energy transfer to the free electron takes place, and the cross section for the scattering is given by the Thomson formula (36.13) of Part I.

As λ approaches Λ (hard X-rays and γ -rays) the classical scattering is replaced by the Compton effect. The change in the wavelength of the light becomes comparable with the wavelength. The radiation knocks out recoil electrons (moving mainly forward in the direction of the incident photon),

and the cross section for scattering, as is seen in fig. I.17, §36, decreases with the energy of the photon. The phenomenon assumes a clearly pronounced quantum character and a classical treatment of the phenomenon of scattering becomes quite impossible.

The quantum-mechanical calculation of the cross section for Compton scattering will be given in Part V.

Relativistic Electrodynamics

§ 18. Charge conservation, the four-dimensional current and the equation of continuity

We shall now turn from relativistic mechanics to relativistic electrodynamics. We shall take as the basis of relativistic electrodynamics the assumption of the invariance and conservation of electric charge. The charge is a fundamental quantity characterizing the properties of particles, and the charge constancy is strictly observed in all known physical processes.

The charge conservation law

$$\nabla \cdot (\rho \mathbf{u}) + \frac{\partial \rho}{\partial t} = 0 \quad (18.1)$$

must hold in all inertial reference frames. In order to give a relativistically invariant form to the charge conservation law, one can, following the usual method, write it in four-dimensional form.

For this it is sufficient to introduce the 4-vector j_α , called the four-dimensional current and defined by the relation

$$j_\alpha = (\rho \mathbf{u}, ic\rho). \quad (18.2)$$

Then (18.1) is easily written in four-dimensional form:

$$\frac{\partial j_\alpha}{\partial x_\alpha} = \frac{\partial j_x}{\partial x} + \frac{\partial j_y}{\partial y} + \frac{\partial j_z}{\partial z} + \frac{\partial j_\tau}{\partial \tau} = 0. \quad (18.3)$$

It is formula (18.3), written in four-dimensional form, which is the relativistically invariant expression. Hence it follows that the quantity j_α , which we determined formally, indeed represents a 4-vector.

From the definition of the 4-vector j_α it follows directly that, in passing from one reference frame to another, its components must transform according to formulae (11.1)–(11.4). If this transformation law is applied to the fourth component $j_\tau = ic\rho$, then we obtain the invariance of the electric charge, $de = \rho dV$ which is present in an arbitrary volume element dV . Indeed, let us consider the change of j_τ in the transition from the reference frame K' in which the charges are at rest to the reference frame K . The reference frame K' moves relative to K with a velocity v . In the reference frame K' the velocity $\mathbf{u} = 0$ and, for the vector j_α , only the component j_τ differs from zero.

From the formula (11.4) for the transformation of the fourth component of the 4-vector we find the expression for the change in the charge density:

$$\rho = \frac{\rho'}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (18.4)$$

Multiplying (18.4) by the volume element dV , we have

$$\rho dV = de = \frac{\rho' dV'}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

The change in the volume (6.2) gives

$$dV = dV' \sqrt{1 - \frac{v^2}{c^2}},$$

so that

$$\rho dV = \rho' dV'.$$

Thus, the charge in any volume element is invariant under Lorentz transformations. This can obviously be interpreted in the following way: as the volume decreases because of the Lorentz contraction the charge density increases in the same ratio, so that the total charge is not changed.

§19. The relativistically invariant formulation of the equations of the electromagnetic field potentials

We have pointed out earlier that the theory of the electromagnetic field was formulated "correctly" from the point of view of the theory of relativity from the very beginning. This means that the Maxwell-Lorentz system of equations is relativistically invariant, satisfying the requirements of the theory of relativity.

One can verify this most simply by considering the equations for the potentials.

According to what was said in §10 of Part I the system of equations

$$\nabla^2 A - \frac{1}{c^2} \frac{\partial^2 A}{\partial t^2} = -\frac{4\pi}{c} \rho u, \quad (19.1)$$

$$\nabla^2 \varphi - \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = -4\pi \rho, \quad (19.2)$$

taking into account the gauge condition

$$\nabla \cdot A + \frac{1}{c} \frac{\partial \varphi}{\partial t} = 0, \quad (19.3)$$

is completely equivalent to the Maxwell-Lorentz equations.

The relativistic invariance of the system (19.1)–(19.3) follows directly from the fact that this system can be written without any changes in the four-dimensional form.

Indeed, we multiply (19.2) by i and note that then the right-hand sides of eqs. (19.1)–(19.2) contain the components of the 4-vector of the current density. Hence it follows that also the left-hand sides represent the components of a 4-vector which we shall call the 4-potential A_α :

$$A_\alpha = (A, i\varphi).$$

By means of the 4-vectors A_α and j_α eqs. (19.1) and (19.2) can be written in the form

$$\square A_\alpha = -\frac{4\pi}{c} j_\alpha, \quad (19.4)$$

where \square denotes D'Alembert's operator (the D'Alembertian):

$$\square = \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} = \frac{\partial^2}{\partial x_\alpha^2}. \quad (19.5)$$

The gauge condition may be written at once in the four-dimensional form

$$\frac{\partial A_\alpha}{\partial x_\alpha} = 0. \quad (19.6)$$

Thus, the complete system of equations for potentials is written in the relativistically invariant form. This means that the laws of electrodynamics are the same in all inertial reference frames.

Further, we see that the electromagnetic field potentials and, consequently, the electric and magnetic field vectors, are not themselves invariant quantities. The relative character of the values of the field vectors is in no way something new and unexpected. It is sufficient to recall that a moving charge produces a magnetic field which, however, is absent in a reference frame moving together with the charge.

The law of transformation of the potentials can easily be written, since they transform according to the general formula for the transformation of the components of a 4-vector (11.1)–(11.4).

We have, obviously,

$$A_x = \frac{A'_x + \frac{v}{c} \varphi'}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (19.7)$$

$$A_y = A'_y, \quad (19.8)$$

$$A_z = A'_z, \quad (19.9)$$

$$\varphi = \frac{\varphi' + \frac{v}{c} A'_x}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (19.10)$$

In the following paragraphs we shall apply these formulae to the treatment of the electromagnetic field of moving charges.

§ 20. The field of a moving charge

Consider the simple problem of finding the electromagnetic field produced by a uniformly moving charge, when the velocity v of the charge is comparable with the velocity of light. In the reference frame K' moving with the charge the magnetic field is absent, and the electric field potential is expressed by the formula

$$\varphi' = e/r'.$$

In the inertial reference frame K with respect to which the charge moves with a velocity v the electric field potential, according to (19.10), has the form

$$\varphi = \frac{\varphi'}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{e}{r' \sqrt{1 - \frac{v^2}{c^2}}}. \quad (20.1)$$

We need the formula for the transformation of the length of the radius vector. We have

$$\begin{aligned} r' &= \sqrt{(x')^2 + (y')^2 + (z')^2} = \sqrt{\left[\frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \right]^2 + y^2 + z^2} = \\ &= \sqrt{\frac{(x - vt)^2 + \left(1 - \frac{v^2}{c^2}\right)(y^2 + z^2)}{1 - \frac{v^2}{c^2}}}. \end{aligned}$$

The asymmetry of this formula is associated with the fact that the motion takes place along the x -axis.

Thus we find

$$\varphi = \frac{e}{\sqrt{(x - vt)^2 + \left(1 - \frac{v^2}{c^2}\right)(y^2 + z^2)}} \quad (20.2)$$

We note that $x = vt$, $y = 0$, $z = 0$ represent the coordinates of the point at which the charge is found at an instant t , so that the radius vector drawn

from the charge to the point of observation N (x, y, z) can be written in the form

$$\mathbf{r} = \mathbf{i}(x - vt) + \mathbf{j}y + \mathbf{k}z \quad (20.3)$$

and

$$r = \sqrt{(x - vt)^2 + y^2 + z^2} \quad (20.4)$$

It is convenient to simplify (20.2), expressing φ in terms of r and the angle ψ formed by the vector \mathbf{r} and the x -axis (fig. 11.5).

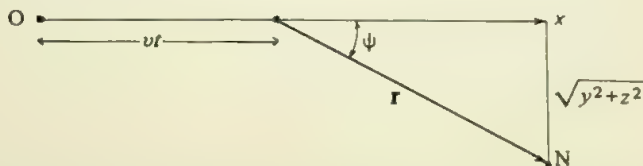


Fig. 11.5

From fig. 11.5 it is seen that

$$x - vt = r \cos \psi,$$

$$y^2 + z^2 = r^2 \sin^2 \psi,$$

and, consequently,

$$\begin{aligned} \sqrt{(x - vt)^2 + \left(1 - \frac{v^2}{c^2}\right)(y^2 + z^2)} &= r \sqrt{\cos^2 \psi + \left(1 - \frac{v^2}{c^2}\right) \sin^2 \psi} = \\ &= r \sqrt{1 - \frac{v^2}{c^2} \sin^2 \psi}. \end{aligned}$$

Hence we have

$$\varphi = \frac{e}{r \sqrt{1 - \frac{v^2}{c^2} \sin^2 \psi}}. \quad (20.5)$$

Passing on to the calculation of the vector potential \mathbf{A} we note that in the reference frame K' there is no magnetic field and $\mathbf{A}' = 0$. From formulae

(19.7) and (19.10) we find

$$\begin{aligned}
 A_x &= \frac{v}{c} \frac{\varphi'}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{v}{c} \varphi = \\
 &= \frac{ev}{c \sqrt{(x - vt)^2 + \left(1 - \frac{v^2}{c^2}\right)(y^2 + z^2)}} = \frac{ev}{cr \sqrt{1 - \frac{v^2}{c^2} \sin^2 \psi}} \\
 A_y &= 0, \quad A_z = 0.
 \end{aligned}$$

In vector form this can be written

$$\mathbf{A} = \varphi \mathbf{v} / c \quad (20.6)$$

Knowing the field potentials one can then find the fields themselves.

In calculating the electric field \mathbf{E} it should be kept in mind that in our case of a uniformly moving charge the differentiation with respect to time reduces to the differentiation with respect to the coordinate x according to the formula

$$\frac{\partial}{\partial t} = -v \frac{\partial}{\partial x}.$$

Indeed, in a uniform motion \mathbf{A} and φ depend on the coordinate and time according to $f(x - vt)$. For such a function

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial (x - vt)} \frac{\partial (x - vt)}{\partial t} = -v \frac{\partial f}{\partial x},$$

from which we find

$$\begin{aligned}
 E_x &= -\frac{\partial \varphi}{\partial x} - \frac{1}{c} \frac{\partial A_x}{\partial t} = -\frac{\partial \varphi}{\partial x} + \frac{v^2}{c^2} \frac{\partial \varphi}{\partial x} = -\left(1 - \frac{v^2}{c^2}\right) \frac{\partial \varphi}{\partial x} = \\
 &= \left(1 - \frac{v^2}{c^2}\right) \frac{e(x - vt)}{\left\{(x - vt)^2 + \left(1 - \frac{v^2}{c^2}\right)(y^2 + z^2)\right\}^{3/2}},
 \end{aligned}$$

$$\begin{aligned}
 E_y &= -\frac{\partial \varphi}{\partial y} - \frac{1}{c} \frac{\partial A_y}{\partial t} = \\
 &= \left(1 - \frac{v^2}{c^2}\right) \frac{ey}{\left\{(x-vt)^2 + \left(1 - \frac{v^2}{c^2}\right)(y^2 + z^2)\right\}^{3/2}}, \\
 E_z &= -\frac{\partial \varphi}{\partial z} - \frac{1}{c} \frac{\partial A_z}{\partial t} = \\
 &= \left(1 - \frac{v^2}{c^2}\right) \frac{ez}{\left\{(x-vt)^2 + \left(1 - \frac{v^2}{c^2}\right)(y^2 + z^2)\right\}^{3/2}},
 \end{aligned}$$

or, in vector form,

$$\mathbf{E} = \left(1 - \frac{v^2}{c^2}\right) \frac{e\mathbf{r}}{r^3 \left(1 - \frac{v^2}{c^2} \sin^2 \psi\right)^{3/2}}. \quad (20.7)$$

When $v \ll c$ formula (20.7) goes over, of course, into the electrostatic expression for the field of a charge at rest.

We can also find the magnetic field \mathbf{H} easily:

$$\mathbf{H} = \nabla \times \mathbf{A} = \frac{1}{c} \nabla \times (\mathbf{v}\varphi) = -\frac{1}{c} \mathbf{v} \times \nabla \varphi = \frac{1}{c} \mathbf{v} \times \mathbf{E}. \quad (20.8)$$

Formulae (20.5) and (20.7) show that, in contrast to the field of a charge at rest, the electric field of a moving charge does not have spherical symmetry.

The scalar potential φ has a constant value at the surface of the ellipsoid:

$$(x-vt)^2 + \left(1 - \frac{v^2}{c^2}\right)(y^2 + z^2) = \text{const.}$$

This ellipsoid is obtained from a sphere compressed in the direction of the x -axis by a factor $\sqrt{1 - v^2/c^2}$.

The character of the field distribution is most clearly seen from the formula

$$|E| = \frac{e \left(1 - \frac{v^2}{c^2}\right)}{r^2 \left(1 - \frac{v^2}{c^2} \sin^2 \psi\right)^{3/2}}. \quad (20.9)$$

On the x -axis ($\psi = 0$) the field is less than the electrostatic field by a factor of $(1 - v^2/c^2)$, while in the plane perpendicular to the x -axis it is increased in the ratio $1/\sqrt{1 - v^2/c^2}$.

As the velocity increases the equipotential ellipsoid gets more and more oblate, while the value of the field in the direction of motion decreases and in the perpendicular direction increases. When $v \sim c$ the field is concentrated in a small angular interval near the plane perpendicular to the direction of motion. The width of this interval is $\Delta\psi \sim \sqrt{1 - v^2/c^2}$.

The magnetic field \mathbf{H} is always perpendicular to the direction of motion and to the electric field vector \mathbf{E} . For $v \sim c$, $|\mathbf{H}| \sim |\mathbf{E}|$. At small velocities ($v \ll c$), it can be assumed that

$$\mathbf{H} = \frac{1}{c} \mathbf{v} \times \mathbf{E} \approx \frac{e}{c} \frac{\mathbf{v} \times \mathbf{r}}{r^3}, \quad (20.10)$$

if it is also assumed that \mathbf{E} has approximately the electrostatic value. This last formula is the same as (20.5) of Part I.

The field produced by a moving charge was found by classical methods * even before the appearance of the theory of relativity. There is nothing surprising in this, since the Maxwell-Lorentz equations are relativistically invariant.

Let us apply the formulae obtained to the calculation of the force of interaction between two charges e_1 and e_2 moving with the same velocity v with respect to the laboratory system. In the reference frame connected with the charges the force is, obviously, equal to $\mathbf{F} = (e_1 e_2 / r^3) \mathbf{r}$ and is directed along the vector connecting the charges. We choose the direction of the vector \mathbf{v} to be the x -axis.

From the point of view of the laboratory system the field of charge e_1 is expressed by formulae (20.8) and (20.10). The charge e_2 is acted upon by the Lorentz force

* See, for example, R. Becker, *Electromagnetic fields and interactions*, Vol. I (Blackie, London, 1964) p. 267.

$$\begin{aligned} \mathbf{F} &= e_2 \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right) = e_2 \mathbf{E} + \frac{e_2}{c^2} \mathbf{v} \times (\mathbf{v} \times \mathbf{E}) = \\ &= e_2 \mathbf{E} + \frac{e_2}{c^2} \mathbf{v} (\mathbf{v} \cdot \mathbf{E}) - \frac{e_2}{c^2} v^2 \mathbf{E} = e_2 \left(1 - \frac{v^2}{c^2} \right) \mathbf{E} + e_2 \frac{\mathbf{v} v E_x}{c^2}. \end{aligned}$$

This force is no longer directed along the radius vector. The component of the force in the direction of motion is equal to

$$F_x = e_2 E_x = \frac{e_1 e_2 \left(1 - \frac{v^2}{c^2} \right) \cos \psi}{r^2 \left(1 - \frac{v^2}{c^2} \sin^2 \psi \right)^{3/2}}.$$

The component of the force in the perpendicular direction is equal to

$$F_y = e_2 \left(1 - \frac{v^2}{c^2} \right) E_y = \frac{e_1 e_2 \left(1 - \frac{v^2}{c^2} \right)^2 \sin \psi}{r^2 \left(1 - \frac{v^2}{c^2} \sin^2 \psi \right)^{3/2}}.$$

Before the appearance of the theory of relativity it was assumed that, by observing the interaction between moving charges, one could determine their absolute velocity with respect to the ether. However, attempts at such measurements did not lead to any positive results.

In the light of the theory of relativity the error in this reasoning is clear: only the common relative velocity \mathbf{v} of the two charges enters into the formula for the force measured in the laboratory system.

It is interesting to apply the result obtained to the motion of a uniformly charged sphere. In the proper reference frame moving together with the sphere the charge density is constant and the lines of field are normal to the surface. In the laboratory system the repulsion between the charges must lead to a non-uniform charge density distribution over the sphere. If the polar axis is drawn in the direction of motion, then the charge density must be largest at the poles of the sphere and a minimum in the equatorial plane.

From the non-relativistic point of view, the absolute velocity of motion of the sphere could be found by measuring this charge distribution. In reality this is not the case: from the point of view of the laboratory system a moving sphere must undergo a contraction and turn into an ellipsoid. Calculation

shows that the effect of contraction of the sphere compensates exactly for the effect of the charge accumulation at the poles. As a result the charge density measured in the laboratory system will be constant over the entire sphere.

This example is very interesting because it shows how the Maxwell-Lorentz theory, "correct" from the point of view of the theory of relativity, leads to false results in association with the classical notions of space and time. Only an alteration of the ideas of space and time in conjunction with the laws of electrodynamics allows one consequently to bring the theory into agreement with experiment.

In addition to the field of a uniformly moving charge it turns out to be possible to find the field of a charge performing an arbitrary motion.

In §25 of Part I we found an expression for the Liénard-Wiechert potentials and pointed out that the corresponding expressions do not lose their applicability even at velocities close to the velocity of light.

To convince ourselves of this, we write the Liénard-Wiechert potentials in the four-dimensional form.

We introduce the 4-vector R_α having the components $(\mathbf{r} - \mathbf{r}_0)$, $ic(t - \tau)$, where \mathbf{r} is the coordinate of the observation point, \mathbf{r}_0 is the coordinate of the charge, and $\tau = t - |\mathbf{r} - \mathbf{r}_0|/c$. The relation between the components of the vector R_α is given by the formula

$$R_\alpha^2 = 0.$$

Indeed, substituting into it the components of R_α , we have

$$(\mathbf{r} - \mathbf{r}_0)^2 - c^2(t - \tau)^2 = 0.$$

By means of the 4-vector R_α one can introduce the Liénard-Wiechert 4-potential by the relation

$$A_\alpha = -eu_\alpha/R_\beta u_\beta.$$

where the summation is carried out over the index β ($\beta = x, y, z, \tau$). Indeed, making use of the definition of the 4-velocity u_β (11.6), we have

$$\begin{aligned}
 R_\beta u_\beta &= R_x u_x + R_y u_y + R_z u_z + R_\tau u_\tau = \\
 &= (x - x_0) \frac{v_x}{\sqrt{1 - \frac{v^2}{c^2}}} + (y - y_0) \frac{v_y}{\sqrt{1 - \frac{v^2}{c^2}}} + \\
 &\quad + (z - z_0) \frac{v_z}{\sqrt{1 - \frac{v^2}{c^2}}} + ic(t - \tau) \frac{ic}{\sqrt{1 - \frac{v^2}{c^2}}} = \\
 &= \frac{c}{\sqrt{1 - \frac{v^2}{c^2}}} \left\{ \frac{(\mathbf{r} - \mathbf{r}_0) \cdot \mathbf{v}}{c} - c(t - \tau) \right\} = \\
 &= \frac{c}{\sqrt{1 - \frac{v^2}{c^2}}} \left\{ \frac{(\mathbf{r} - \mathbf{r}_0) \cdot \mathbf{v}}{c} - |\mathbf{r} - \mathbf{r}_0| \right\} = \\
 &= \frac{c}{\sqrt{1 - \frac{v^2}{c^2}}} \left\{ \frac{\mathbf{R}(\tau) \cdot \mathbf{v}}{c} - R(\tau) \right\},
 \end{aligned}$$

where, in correspondence with the notation of § 25 of Part I, $\mathbf{R}(\tau) = \mathbf{r} - \mathbf{r}_0$ is the radius vector drawn from an instantaneous position of the charge to the observation point at an instant τ .

Correspondingly, the components of A_α have the form

$$\begin{aligned}
 A_x &= - \frac{ev_x}{\sqrt{1 - \frac{v^2}{c^2}}} \frac{1}{\frac{c}{\sqrt{1 - \frac{v^2}{c^2}}} \left\{ \frac{\mathbf{R}(\tau) \cdot \mathbf{v}}{c} - R(\tau) \right\}} = \\
 &= \frac{ev_x}{c \left\{ R(\tau) - \frac{\mathbf{R}(\tau) \cdot \mathbf{v}}{c} \right\}} \quad (20.11)
 \end{aligned}$$

with analogous expressions for A_y and A_z ;

$$\begin{aligned}
 A_\tau = i\varphi &= \frac{iec}{\sqrt{1-\frac{v^2}{c^2}}} \frac{1}{\frac{c}{\sqrt{1-\frac{v^2}{c^2}}} \left\{ \frac{\mathbf{R}(\tau) \cdot \mathbf{v}}{c} - R(\tau) \right\}} = \\
 &= \frac{ie}{\left\{ R(\tau) - \frac{\mathbf{R}(\tau) \cdot \mathbf{v}}{c} \right\}}
 \end{aligned}$$

Hence

$$\varphi = \frac{e}{\left\{ R(\tau) - \frac{\mathbf{R}(\tau) \cdot \mathbf{v}}{c} \right\}} \quad (20.12)$$

Formulae (20.11) and (20.12) are identical with formulae (25.3) and (25.4) of Part I.

Thus, the Liénard-Wiechert potentials can be written in relativistically invariant form and, consequently, use can be made of them at an arbitrary value of the velocity of motion of the charge.

§21. The electromagnetic field tensor and Maxwell's equations

We now turn to the calculation of the components of the electric and magnetic fields.

Making use of the definition of the 4-potential and introducing the coordinates of the radius 4-vector, we can write the components of the electric field in the form

$$\begin{aligned}
 E_x &= i \left(\frac{\partial A_\tau}{\partial x} - \frac{\partial A_x}{\partial \tau} \right), \\
 E_y &= i \left(\frac{\partial A_\tau}{\partial y} - \frac{\partial A_y}{\partial \tau} \right), \\
 E_z &= i \left(\frac{\partial A_\tau}{\partial z} - \frac{\partial A_z}{\partial \tau} \right),
 \end{aligned} \quad (21.1)$$

while the components of the magnetic field are expressed in terms of the components of the vector potential by the usual relations

$$H_x = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right), \quad (21.2)$$

and analogously for H_y and H_z .

The symmetry of formulae (21.1) and (21.2) suggests that we try to write the whole set in the form of one general formula. To do this we introduce a tensor $F_{\alpha\beta}$ by means of the relation

$$F_{\alpha\beta} = \frac{\partial A_\beta}{\partial x_\alpha} - \frac{\partial A_\alpha}{\partial x_\beta}. \quad (21.3)$$

The tensor $F_{\alpha\beta}$ is antisymmetric by definition. The calculation of the components of $F_{\alpha\beta}$ leads to the result

$$F_{\alpha\beta} = \begin{vmatrix} 0 & H_z & -H_y & -iE_x \\ -H_z & 0 & H_x & -iE_y \\ H_y & -H_x & 0 & -iE_z \\ iE_x & iE_y & iE_z & 0 \end{vmatrix}. \quad (21.4)$$

We see that all components of the electric and magnetic field vectors prove to be components of one tensor quantity $F_{\alpha\beta}$. It turns out that Maxwell's equations represent a system of equations for the tensor $F_{\alpha\beta}$. That is, if we write for $F_{\alpha\beta}$ the equation

$$\frac{\partial F_{\alpha\beta}}{\partial x_\lambda} + \frac{\partial F_{\beta\lambda}}{\partial x_\alpha} + \frac{\partial F_{\lambda\alpha}}{\partial x_\beta} = 0, \quad (21.5)$$

then, assuming α, β and λ to be successively equal to 1, 2, 3 and 4 and making use of the definition (21.4), we find easily that the four-dimensional equation (21.5) represents a notation of the two Maxwell-Lorentz vector equations (8.1) and (8.2) in Part I.

Analogously, writing the four-dimensional equation

$$\frac{\partial F_{\alpha\beta}}{\partial x_\beta} = \frac{4\pi}{c} j_\alpha, \quad (21.6)$$

we can verify that it involves the two vector equations (8.3) and (8.4). Formulae (21.5) and (21.6) represent the relativistically invariant form of notation of the Maxwell-Lorentz system of equations. The indissoluble connection between the electric and magnetic fields appeared in the non-relativistic formulation of the theory of the electromagnetic field as a proposition which followed from the whole of the experimental data. In relativistic electrodynamics this connection proves to be inevitable and self-evident. From the very definition of the field it follows that its properties are characterized not by two vectors but by one antisymmetric tensor. The equations of the electromagnetic field are equations related to this tensor. The components of the fields \mathbf{E} and \mathbf{H} appear equivalently in the tensor $F_{\alpha\beta}$. Under Lorentz transformations, which we can now formulate, the fields \mathbf{E} and \mathbf{H} are expressed in terms of each other.

Indeed, according to formulae (11.21) we have (assuming $A_{xy} = H_z$, $A_{xz} = -H_y$, $A_{yz} = H_x$, $A_{x\tau} = -iE_x$, $A_{y\tau} = -iE_y$, $A_{z\tau} = -iE_z$):

$$E_x = E'_x, \quad E_y = \frac{E'_y - \frac{v}{c}H'_z}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad E_z = \frac{E'_z + \frac{v}{c}H'_y}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (21.7)$$

$$H_x = H'_x, \quad H_y = \frac{H'_y + \frac{v}{c}E'_z}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad H_z = \frac{H'_z - \frac{v}{c}E'_y}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (21.8)$$

We see that statements of the type "the field is of purely electric character" or "the field is of purely magnetic character" are relative. An electric or magnetic field can be equal to zero in one reference frame and different from zero in another. Hence it makes no sense to attribute physical reality to the electric field and magnetic field separately. It is their combination, expressed by the electromagnetic field tensor $F_{\alpha\beta}$, which is the physical reality.

It is natural to establish which quantities characterizing the electromagnetic field are invariant. Since the electromagnetic field is described by the antisymmetric tensor $F_{\alpha\beta}$ then according to the results presented in §11 these invariants are the quantities $F_{\alpha\beta}F_{\alpha\beta}$ and $F_{\alpha\lambda}F_{\lambda\mu}F_{\mu\nu}F_{\nu\alpha}$. The cube invariant $F_{\alpha\beta}F_{\beta\lambda}F_{\lambda\alpha}$ turns out to be equal to zero. A simple calculation gives

$$I_1 = F_{\alpha\beta} F_{\alpha\beta} = H^2 - E^2 = \text{invar} , \quad (21.9)$$

$$I_2 = F_{\alpha\lambda} F_{\lambda\mu} F_{\mu\nu} F_{\nu\alpha} = (\mathbf{E} \cdot \mathbf{H})^2 = \text{invar} . \quad (21.10)$$

The invariants I_1 and I_2 are the absolute characteristics of the field. The statement "the electromagnetic field is equal to zero" ($I_1 = I_2 = 0$) or "the electric and magnetic field have the same value and are perpendicular to each other" ($I_1 = I_2 = 0$) are examples of absolute statements.

It is also true that, if $I_2 = 0$, i.e. if the fields \mathbf{E} and \mathbf{H} are perpendicular to each other, then they will remain perpendicular to each other in all reference frames. If in this case $I_1 > 0$, then in all reference frames $|\mathbf{H}| > |\mathbf{E}|$. We can in such cases find a reference frame in which the electric (but not the magnetic) field is equal to zero. Analogously, if $I_1 < 0$, one can find a reference frame in which the magnetic field is absent.

In the limiting case $v \ll c$ the Lorentz transformation formulae are essentially simplified. Disregarding v^2/c^2 , they can be written in the form

$$\mathbf{E} = \mathbf{E}' - \frac{\mathbf{v} \times \mathbf{H}'}{c} , \quad \mathbf{H} = \mathbf{H}' + \frac{\mathbf{v} \times \mathbf{E}'}{c} . \quad (21.11)$$

These transformations had already been obtained in non-relativistic field theory.

In conclusion we note that the relation and resemblance between the fields \mathbf{E} and \mathbf{H} does not at the same time exclude the existence of an essential difference between them, which we have mentioned in §7 of Part I. In relativistic electrodynamics this difference manifests itself in the fact that the components of the electric field are the temporal (imaginary) components of the electromagnetic field tensor, whereas the components of the magnetic field form the set of its spatial (real) components.

§22. Some applications: Doppler effect, Mössbauer effect, observation of rapidly moving bodies, transformation of angles, intensities and cross sections

We have seen that the statement "the electromagnetic field is absent at a certain point of space at a particular instant of time" has an absolute character. From this it follows that the value of the phase α in the electromagnetic

$$\mathbf{E} = \mathbf{E}_0 e^{i\alpha} , \quad \mathbf{H} = \mathbf{H}_0 e^{i\alpha} ,$$

wave is an invariant. If, for example, the phase is equal to $\frac{1}{2}\pi$ or to an integer multiple of $\frac{1}{2}\pi$, so that $\mathbf{E} = \mathbf{H} = 0$, then this value of the phase must be conserved in all inertial reference frames.

Formally introducing the 4-vector k_α ,

$$k_\alpha = \left(\mathbf{k}, i \frac{\omega}{c} \right),$$

called the four-dimensional wave vector, one can write the phase of the wave in the form of the scalar product of two 4-vectors, k_α and r_α ,

$$\alpha = k_\alpha r_\alpha = (\mathbf{k} \cdot \mathbf{r} - \omega t).$$

Since the phase is invariant, the above formula shows that the quantity k_α , determined formally, is clearly a 4-vector.

The law of transformation of the components of a 4-vector also allows one to find the law of transformation of frequencies in the theory of relativity. That is to say, from the definition of the four-dimensional wave vector it follows that its fourth component transforms according to the law

$$k_r = \frac{k'_r + i \frac{v}{c} k'_x}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Expressing k_r in terms of the frequency ω , we find

$$\omega = \frac{\omega' + vk'_x}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Writing k'_x in the form $k'_x = (\omega'/c) \cos \theta'$, where $\cos \theta'$ is the direction cosine of the wave vector, we find

$$\omega = \frac{\omega' \left(1 + \frac{v}{c} \cos \theta' \right)}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (22.1)$$

Analogously,

$$k_x = \frac{\omega}{c} n_x = \frac{k'_x - i k'_y \frac{v}{c}}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{\frac{\omega'}{c} \cos \theta' + \frac{v}{c^2} \omega'}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (22.2)$$

Expressing ω in terms of ω' , we find

$$\cos \theta = \frac{\cos \theta' + \frac{v}{c}}{1 + \frac{v}{c} \cos \theta'}, \quad (22.3)$$

$$\sin \theta = \sin \theta' \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 + \frac{v}{c} \cos \theta'},$$

The formulae for the inverse transformation from the reference frame K to the reference frame K' read:

$$\omega' = \frac{\omega \left(1 - \frac{v}{c} \cos \theta\right)}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (22.4)$$

$$\cos \theta' = \frac{\cos \theta - \frac{v}{c}}{1 - \frac{v}{c} \cos \theta}, \quad (22.5)$$

$$\sin \theta' = \sin \theta \frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 - \frac{v}{c} \cos \theta}.$$

Formula (22.1) expresses the Doppler effect in the theory of relativity. As is well known, the Doppler effect consists in a change of the frequency emitted by a moving source in comparison with the frequency emitted by the same source at rest.

Formulae (22.3) and (22.5) give the law of angular transformation. The law of angular transformation given by these relations is the same as that following from (7.4). One can verify this most simply by dividing the lower formula in (22.3) by the upper one.

In these formulae ω' is the frequency measured in the reference frame moving together with the source of radiation, and ω is the frequency in the reference frame at rest. It is assumed that the source moves together with the reference frame K' along the x -axis. The angle θ' represents the angle between the direction of the emitted wave, characterized in the reference frame K' by a vector \mathbf{k}' , and the direction of motion of the source. In the transition to the reference frame at rest the angle θ' transforms into the angle θ between the vector \mathbf{k} and the direction of motion.

Formula (22.4), which can be written in the form

$$\omega = \frac{\omega' \sqrt{1 - \frac{v^2}{c^2}}}{1 - \frac{v}{c} \cos \theta}, \quad (22.6)$$

is the most important for practical use. Formula (22.6) allows one to find the frequency of light ω as a function of the frequency of light ω' emitted by the moving source in the proper reference frame K' and the angle θ measured in the reference frame K .

If the source of light approaches or moves away from the observer who is located, for example, at the origin of the reference frame K , then one speaks about the longitudinal Doppler effect. For the approaching case $\cos \theta = 1$ and the frequency $\omega > \omega'$, and for the receding case $\cos \theta = -1$ and the frequency $\omega < \omega'$.

It is interesting to compare the relativistic formula for the change in the frequency in the Doppler effect with the classical one. The latter is obtained from the elementary treatment of waves emitted by the moving source and reaching the observer, and has the form

$$\omega = \omega' \left(1 + \frac{v}{c} \cos \theta \right). \quad (22.7)$$

The value of the angle was assumed in classical physics to be invariant in all reference frames.

Comparing the classical and relativistic formula for the change in the frequency in the Doppler effect we see that the two formulae are the same when

the source of radiation is moving with a velocity $v \ll c$, so that it can be assumed that

$$\sqrt{1 - \frac{v^2}{c^2}} \approx 1$$

and

$$\frac{1}{1 - \frac{v}{c} \cos \theta} \approx 1 + \frac{v}{c} \cos \theta.$$

However, if the second order terms are taken into account, then an essential difference arises between these formulae. In particular, when the source is moving in a direction perpendicular to the direction of observation, i.e. when $\theta = \frac{1}{2}\pi$, the classical formula (22.7) shows that no change in the frequency takes place. On the contrary, according to the theory of relativity a change in the frequency does take place (the "transverse" Doppler effect) which is, according to (22.6), equal to

$$\omega = \omega' \sqrt{1 - \frac{v^2}{c^2}}. \quad (22.8)$$

The experimental study of the Doppler effect was of particularly great importance because the change in the frequency is connected directly with the change in time in the transition from one inertial reference frame to another. The experimental investigation of the Doppler effect made it possible to confirm to a high degree of accuracy the validity of the relativistic relations.

In Ives' experiments* the change in the frequency emitted by hydrogen atoms in canal rays was investigated. The velocity of the atoms was about $6 \times 10^{-3} c$. The main difficulty lay in the separation of the second order effect with respect to (v/c) , which is small in comparison with the ordinary (classical) line shift by an amount of the order of magnitude of v/c . To do this light rays emitted along and against the direction of motion were matched by means of mirrors. According to the relativistic formula (22.1) the value of the mean of the emitted frequencies of the lines is equal to

* Ives' and Stillwell's. See W. Pauli, *Theory of relativity* (Pergamon, London, 1958).

$$\begin{aligned}
 \bar{\omega} &= \frac{\omega_1 + \omega_2}{2} = \frac{\omega' \left(1 - \frac{v}{c}\right)}{2\sqrt{1 - \frac{v^2}{c^2}}} + \frac{\omega' \left(1 + \frac{v}{c}\right)}{2\sqrt{1 - \frac{v^2}{c^2}}} = \\
 &= \frac{\omega'}{\sqrt{1 - \frac{v^2}{c^2}}} \approx \omega' \left[1 + \frac{1}{2} \left(\frac{v}{c}\right)^2\right]. \quad (22.9)
 \end{aligned}$$

The classical result would be $\bar{\omega} = \omega'$.

Measurements confirmed to a high degree of accuracy the relativistic formula (22.9).

Ives' experiments, dating from 1938, appeared to be the first direct experimental confirmation of the relativistic law of change of time under Lorentz transformations. From this point of view they played a role analogous to Michelson's experiments. Taking these two experiments as a basis, one could have constructed the entire scheme of the theory of relativity, if both these experiments had preceded the development of the theory.

An extremely accurate measurement of the second order Doppler effect has recently become possible through the discovery of an important new phenomenon, the so-called Mössbauer effect. It is well known that γ -quanta emitted by nuclei are monochromatic to a high degree. The width of the corresponding spectral lines is smaller, by many orders of magnitude, than the natural width of the lines emitted by atoms in optical transitions. This fact was for a long time an obstacle to the observation of resonance absorption by nuclei, i.e. the absorption of γ -quanta of natural frequencies by the nuclei.

Nuclei, as well as atomic systems, absorb radiation of the same frequency which they themselves emit, very strongly. However, in contrast to emitting atomic systems, nuclei undergo a considerable recoil in emitting γ -quanta. This recoil changes the frequency of the emitted γ -quantum, which, owing to the very small width of the lines, leads to a complete shift from the resonance frequency.

Mössbauer showed that the situation is fundamentally changed if the nuclei emitting γ -quanta are in a crystal lattice. The forces of interaction between atoms and their neighbours in a crystal lattice is very large. Hence the recoil momentum acquired by the nucleus in emitting a γ -quantum is insufficient to pull the nucleus out of its position in the lattice. The recoil

momentum is transferred to the crystal as a whole. * The mass of the latter is very large and the emission of a γ -quantum takes place practically without a recoil. Gamma-rays emitted by nuclei in a crystal lattice have an unshifted frequency ν_0 . When γ -rays emitted by a crystal emitter without recoil are passed through an absorber containing the same nuclei as the emitter a resonance absorption is observed.

Although the emission of γ -quanta takes place without a recoil and the momentum of the crystal can be assumed to be equal to zero, the presence of thermal motion leads to a small change in the frequency. Namely, since the velocity of the emitter — an atomic nucleus performing thermal motion — differs from zero, a Doppler effect must occur.

To calculate the change in the frequency use can be made of the following.

If the nucleus emits radiation of frequency ν_0 , then the energy of the γ -quantum is equal to $h\nu_0$. The mass of the γ -quantum is equal to $h\nu_0/c^2$. As a result of the emission the mass of the nucleus is reduced by this amount. Let the mass of the nucleus before emission be equal to M . Then its kinetic energy is

$$E_{\text{kin}}^{(0)} = \sqrt{p^2 c^2 + M^2 c^4} - Mc^2 \approx p^2 / 2M.$$

In emission the mass of the nucleus is reduced by an amount $h\nu_0/c^2$ to

$$M' = M - \frac{h\nu_0}{c^2}$$

for a constant value of the momentum. Consequently the kinetic energy changes by an amount

$$\begin{aligned} \Delta E_{\text{kin}} &= (\sqrt{p^2 c^2 + M'^2 c^4} - M' c^2) - E_{\text{kin}}^{(0)} \approx \\ &\approx \frac{p^2}{2M'} - \frac{p^2}{2M} \approx \frac{p^2}{2M^2} \frac{h\nu_0}{c^2}. \end{aligned}$$

Replacing the true value p^2 by the mean value $M^2 \overline{v^2}$ for thermal motion, we obtain

* At least at a sufficiently low temperature. For more details about the conditions under which the momentum is transferred to the crystal as a whole see the articles: R.Mössbauer, Soviet Physics Uspekhi 3 (1961) 866; and F.L.Shapiro, Soviet Physics Uspekhi 3 (1961) 881.

$$\Delta E_{\text{kin}} \approx h\nu_0 \frac{\overline{v^2}}{2c^2}.$$

The increase in the kinetic energy of the moving emitting nucleus means that the γ -quantum takes away an energy $h\nu_0 - \Delta E_{\text{kin}} = h\nu_0(1 - \overline{v^2}/2c^2)$. This energy is smaller by an amount ΔE_{kin} than the energy of the γ -quantum emitted by the source at rest.

Thus, the thermal motion of emitting nuclei leads to a frequency shift from ν_0 to ν such that

$$\nu = \nu_0 \left(1 - \frac{\overline{v^2}}{2c^2}\right).$$

This expression for the shifted frequency is the same as the formula for the second order Doppler effect (taking into account terms of the order of v^2/c^2). The absence of the linear term in the formula for the Doppler effect is associated with the fact that the mean velocity of thermal motion is equal to zero.

The calculated shift of the emitted line by an amount of $\nu_0 \overline{v^2}/2c^2$ is in good agreement with experimental data.

Since the mean square velocity of a molecule in a crystal depends on the temperature, one can find the temperature dependence of the value of the shift $\nu - \nu_0$:

$$\frac{d\Delta\omega}{dT} = -\frac{\omega_0}{Mc^2} \frac{d}{dT} \frac{M\overline{v^2}}{2}.$$

But in a crystal the heat capacity at a temperature substantially higher than the Debye temperature is equal to (see (30.1) of Part III):

$$C_V = N \frac{d\bar{\epsilon}}{dT} = N \frac{d}{dT} \left(\frac{M\overline{v^2}}{2} + \frac{M\omega^2 \overline{q^2}}{2} \right) = N \frac{dM\overline{v^2}}{dT},$$

hence

$$\frac{d\Delta\omega}{dT} = -\frac{\omega_0}{2Mc^2N} C_V$$

The accuracy of the measurements using the Mössbauer effect is so high that this effect has also been measured. We cannot dwell here on a number of



Other very important results obtained in studying the Mössbauer effect. *

We shall now consider the formulae for the transformation of some quantities which play important roles in physics. We begin with the consideration of the radiation emitted by a particle moving with a velocity close to the velocity of light.

We assume that in the reference frame moving with the particle the emission can take place at a large angle to the direction of motion, i.e. $\cos \theta'$ varies in the range

$$0 \leq \cos \theta' \leq 1.$$

Let us find the angle at which the radiation will be emitted in the motionless (laboratory) reference frame.

It is convenient to transform with formula (22.3). Setting

$$v \approx c - \frac{c}{\alpha^2},$$

where

$$\alpha = - \frac{\sqrt{2}}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

For v close to c the quantity $\alpha^2 \gg 1$. Introducing this notion, we obtain for $\alpha \gg 1$:

$$\begin{aligned} \cos \theta &= \frac{\cos \theta' + 1 - \frac{1}{\alpha^2}}{1 + \left(1 - \frac{1}{\alpha^2}\right) \cos \theta'} = \frac{1 - \frac{1}{\alpha^2(1 + \cos \theta')}}{1 - \frac{\cos \theta'}{\alpha^2(1 + \cos \theta')}} \approx \\ &\approx 1 + \frac{\cos \theta' - 1}{\cos \theta' + 1} \frac{1}{\alpha^2} = 1 - \frac{\tan^2 \frac{1}{2} \theta'}{\alpha^2}. \end{aligned}$$

* See F.L.Shapiro, Soviet Physics Uspekhi 3 (1961) §81; and also the articles by R.Mössbauer and R.V.Pound in the same volume.

Since $\alpha^2 \gg 1$, $\cos \theta$ is close to unity and the angle θ itself is given approximately by

$$\theta \sim \frac{\sqrt{2} \tan \frac{1}{2} \theta'}{\alpha} \sim \tan \frac{1}{2} \theta' \sqrt{1 - \frac{v^2}{c^2}},$$

and is very small for a sufficiently large value of α , i.e. for a small value of the difference $c - v$. This means that in the reference frame at rest the overall radiation is directed forward and is confined to a narrow cone of angle $\Delta \theta \sim \sqrt{1 - v^2/c^2}$. The aperture of this cone is smaller the closer the velocity v of the source is to the velocity of light.

We now obtain the formula for the transformation of an element of solid angle $d\Omega$, which is important for many applications. We have, by definition,

$$d\Omega = 2\pi \sin \theta \, d\theta = -2\pi \, d \cos \theta,$$

$$d\Omega' = -2\pi \, d \cos \theta'.$$

Hence

$$d\Omega = \frac{d \cos \theta}{d \cos \theta'} \, d\Omega'. \quad (22.10)$$

From formula (22.5) we find

$$1 + \frac{v}{c} \cos \theta' = \left(1 - \frac{v^2}{c^2}\right) \frac{1}{1 - \frac{v}{c} \cos \theta}. \quad (22.11)$$

Differentiating (22.11) with respect to $\cos \theta$, we find

$$\frac{d \cos \theta'}{d \cos \theta} = \frac{1 - \frac{v^2}{c^2}}{\left(1 - \frac{v}{c} \cos \theta\right)^2}$$

thus we finally obtain

$$d\Omega' = \frac{1 - \frac{v^2}{c^2}}{\left(1 - \frac{v}{c} \cos \theta\right)^2} d\Omega. \quad (22.12)$$

Let us find how to transform the energy in an electromagnetic wave. From the general relation (13.9) and formula (33.12) of Part I we have

$$E = \frac{E' + \frac{E'}{c} (\mathbf{n}' \cdot \mathbf{v})}{\sqrt{1 - \frac{v^2}{c^2}}} = E' \frac{1 + \frac{v}{c} \cos \theta'}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (22.13)$$

By virtue of the relation (27.10) of Part I the invariant of the energy-momentum 4-vector for the electromagnetic field is equal to zero:

$$\text{invariant} = g^2 - \frac{E^2}{c^2} = 0. \quad (22.14)$$

We note that, writing down the formulae for the energy and momentum of the field, which were determined in Part I as the densities of the corresponding quantities, they are related to unit volume moving with the field, i.e. with the velocity of light. We should therefore define more precisely what such a volume means in the theory of relativity. *

Comparing (22.13) with formula (22.1) for the transformation of frequency, we arrive at the important equality:

$$\frac{E}{\omega} = \frac{E'}{\omega'} = \text{invar}. \quad (22.15)$$

We now find the expression for the transformation of the total power of the radiation:

$$I = - \frac{dE}{dt}.$$

Since the total momentum emitted by the system in that reference frame in which it was at rest at the moment of emission is equal to zero (see (28.5) of Part I), one can write

* This transformation can be found, for instance, in the book: W. Pauli, *Theory of relativity* (Pergamon, London, 1958), or in the repeatedly quoted book by R. Becker.

$$E = \frac{E'}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

ence using (6.3)

$$\frac{dE}{dt} = \frac{d}{dt} \frac{E'}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{dE'}{\sqrt{1 - \frac{v^2}{c^2}}} \frac{\sqrt{1 - \frac{v^2}{c^2}}}{dt'} = \frac{dE'}{dt'},$$

and, consequently,

$$I = - \frac{dE}{dt} = - \frac{dE'}{dt'} = \text{invar}. \quad (22.16)$$

Thus, the total power radiated (the energy emitted per unit time) is an invariant. The radiated power in a solid angle element $d\Omega$ transforms according to the formula

$$dI = I(\theta) d\Omega = I'(\theta') d\Omega' = \text{invar},$$

from which it follows that

$$I(\theta) = I'(\theta') \frac{d\Omega'}{d\Omega} = I'(\theta') \frac{1 - \frac{v^2}{c^2}}{\left(1 - \frac{v}{c} \cos \theta\right)^2}. \quad (22.17)$$

Combining formulae (22.4) and (22.12) we also obtain the expression

$$\omega^2 d\Omega = \omega'^2 d\Omega' = \text{invar} \quad (22.18)$$

In practice one often has to consider the law of transformation of the quantities dk_x, dk_y, dk_z . From (22.2) we find

$$dk'_x = \frac{dk'_x + \frac{v d\omega'}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{dk'_x \left(1 + \frac{v}{c^2} \frac{d\omega'}{dk'_x}\right)}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Taking into account that $\omega' = c\sqrt{k_x'^2 + k_y'^2 + k_z'^2}$, we have

$$\frac{d\omega'}{dk'_x} = \frac{k'_x c^2}{\omega'}$$

so that

$$dk'_x = \frac{dk'_x \left(1 + \frac{vk'_x}{\omega'}\right)}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{dk'_x}{\omega'} \frac{\omega' + vk'_x}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{dk'_x}{\omega'} \omega'.$$

In addition

$$dk'_y = dk_y, \quad dk'_z = dk_z,$$

whence we find

$$\frac{dk_x dk_y dk_z}{\omega} = \frac{dk'_x dk'_y dk'_z}{\omega'} = \text{invar}, \quad (22.19)$$

or, in spherical coordinates,

$$\frac{k^2 dk d\Omega}{\omega} = \frac{k'^2 dk' d\Omega'}{\omega'} = \text{invar}. \quad (22.20)$$

An important problem in the contemporary physics of high-energy particles is that of transformation of the total scattering cross section in the transition from one inertial reference frame to another.

In §43 of Part I we have seen that the total cross section in classical physics is an invariant. In the theory of relativity the cross section in general is not an invariant.

According to (43.2) and (43.4) of Part I the total number of particles scat-

tered by a volume V in a time dt in the reference frame K in which the scatterer is at rest is equal to

$$dN = \sigma I_0 \rho V dt,$$

where ρ is the particle density of the scatterer, and

$$I_0 = n |\mathbf{v}_2 - \mathbf{v}_1| = n v_{\text{rel}},$$

and n is the density of particles in the beam to be scattered, \mathbf{v}_2 is the velocity of particles to be scattered, and \mathbf{v}_1 is the velocity of the particles of the target, equal to zero in the reference frame K .

In a reference frame K' let the target move with a velocity \mathbf{v}_1 . Then in view of the invariance of the number of particles dN we have

$$dN = \sigma I_0 \rho V dt = \sigma' V' I_0' \rho' dt',$$

where

$$I_0' = n' |\mathbf{v}_2' - \mathbf{v}_1'|.$$

Hence it follows that

$$\sigma n \rho |\mathbf{v}_2 - \mathbf{v}_1| V dt = \sigma' n' \rho' |\mathbf{v}_2' - \mathbf{v}_1'| V' dt'.$$

From formulae (6.2) and (6.3) the product $V dt$ is an invariant. The density of the incident beam of particles n and the density of the particles of the target ρ satisfy the continuity equation. In accordance with this, we can construct the 4-vectors $(\rho \mathbf{v}_1, i c \rho)$ and $(n \mathbf{v}_2, i c n)$. We form their scalar product, which is an invariant. We have, obviously

$$(\rho \mathbf{v}_1, i c \rho) \cdot (n \mathbf{v}_2, i c n) = n' \rho' \mathbf{v}_2' \cdot \mathbf{v}_1' - c^2 \rho' n' = -c^2 \rho n,$$

since in the reference frame K the target is at rest and $\mathbf{v}_1 = 0$.

On transforming we have

$$\rho' n' \left(1 - \frac{\mathbf{v}_1' \cdot \mathbf{v}_2'}{c^2} \right) = \rho n.$$

Thus, we can write that

$$\sigma |v_2 - v_1| = \sigma v_{rel} = \frac{\sigma' |v'_2 - v'_1|}{1 - \frac{v'_1 \cdot v'_2}{c^2}}.$$

Since $v_1 = 0$, the last formula can be written in symmetric form:

$$\frac{\sigma |v_2 - v_1|}{1 - \frac{v_1 \cdot v_2}{c^2}} = \frac{\sigma' |v'_2 - v'_1|}{1 - \frac{v'_1 \cdot v'_2}{c^2}}. \quad (22.21)$$

We see that the total scattering cross section is not an invariant.

In the particular case when the velocities of the incident particles and the target are in the same direction or in opposite directions we can choose the direction of motion as an axis and make use of the formula for the addition of velocities. According to the inverse of (7.1) we have

$$v_2 = \frac{v'_2 - v'_1}{1 - \frac{v'_1 v'_2}{c^2}}.$$

Thus, in this particular case

$$\sigma = \sigma' = \text{invar}. \quad (22.22)$$

A question of interest is, what is the form in which rapidly moving objects appear when they are recorded on a photographic plate or when they are observed visually? In other words, will the Lorentz contraction be noticeable? For example, will a sphere be seen in the form of an ellipsoid, and a cube in the form of a parallelepiped, and so on? The corresponding investigation has recently been carried out by Terrel *. It turns out that rapidly moving objects are seen not as flattened but as having undergone a rotation through some angle depending on the velocity of the object with respect to the observer and the angle of observation. In order to understand this, at first sight unexpected, result, we shall define more precisely the difference between the results of a measurement of the form of a rapidly moving body and the special case of recording this form on a photographic plate. In general, to measure the form of a body, for example by recording the radiation appearing

* J. Terrel, Phys. Rev. 116 (1950) 1041. See also V. Weisskopf, Physics Today 13 (1960) 24.

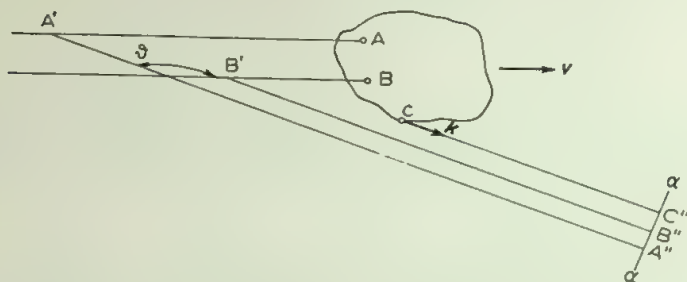


Fig. II.6

simultaneously from different points of its surface, we have to take into account the finite value of the velocity of light.

Different points of the object of observation — a moving body — are located at different distances from the recording device. Hence electromagnetic waves emitted by different points of the surface of the body traverse different paths for different times before arriving at the device. To obtain the true form of the object of observation it is necessary to introduce corresponding corrections into the measurement data.

The situation is different in photographing or observing a body visually. A photographic plate (or an eye) records the radiation reaching it at a given instant. Consequently, a photographic plate simultaneously records electromagnetic waves emitted by different points of the object of observation at different times. Let a photographic plate $\alpha\alpha$ be at rest (fig. II.6) in a certain reference frame. We assume that an extended body is moving in a direction towards it. We denote by v the velocity of the body (and the reference frame K' connected with it). The angle ϑ between a straight line drawn from the photographic plate to the body and the velocity vector v will be called the observation angle.

The radiation from the body can be characterized by a wave vector k' in the reference frame K' connected with the body. The observation angle in the reference frame K' will be equal to ϑ' . The angle θ' , formed by the wave vector k' with the velocity v is obviously equal to $\theta' = \pi - \vartheta'$.

In the reference frame K the wave vector forms an angle $\theta = \pi - \vartheta$ with the vector v . We assume that the solid angle subtended by the object of observation is sufficiently small. The radiation emitted by different points of the object can then be characterized by one value of the wave vector k . The image of the body on the photographic plate obtained at a certain instant will be called "the picture". Let us see how the picture is obtained, i.e. how elec-

tromagnetic waves emitted by different points A, B, C of the surface of the body are recorded (see fig. II.6). The waves emitted by the points A and B at the instant when they were respectively in the positions A' and B' reach the plate at the same time as the wave emitted by the point C on the surface of the body. In other words, radiation emitted by the points A, B, C is recorded simultaneously on the photographic plate at the points A'', B'', C''. However, the waves coming from the more distant points A and B were emitted at earlier times. Let us find the relation between the observation angles ϑ and ϑ' .

Taking into account the above mentioned relation between the observation angles ϑ , ϑ' and the emission angles θ , θ' and formula (22.5), we can write

$$\sin \vartheta' = \frac{\sqrt{1 - \frac{v^2}{c^2}} \sin \vartheta}{1 + \frac{v \cos \vartheta}{c}}. \quad (22.23)$$

In formula (22.5) we have replaced the angles θ and θ' by $\pi - \vartheta$, $\pi - \vartheta'$. The dependence $\vartheta' = f(\vartheta)$ is presented graphically in fig. II.7. The dotted line in the same drawing shows the straight line $\vartheta' = \vartheta$, corresponding to the case $v = 0$, i.e. to the recording on the photographic plate of an object which is at rest with respect to it. Consider first the case $\vartheta = \pi$. This means that the body to be photographed is moving directly towards the photographic plate. In this case $\vartheta' = \pi$, i.e. we obtain the image of the front surface of the body on the photographic plate.

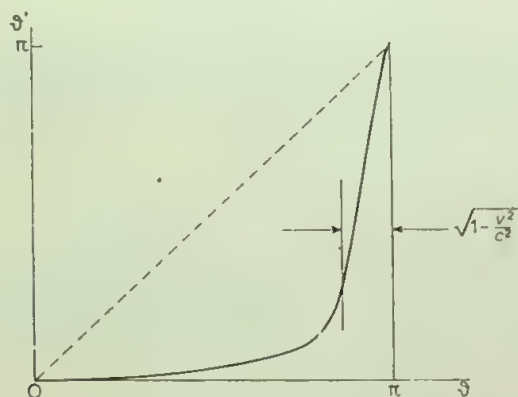


Fig. II.7

Now let the angle ϑ have a value smaller than $\pi - \sqrt{1 - v^2/c^2}$. This means that the observation angle lies outside a narrow cone about the direction of motion.

From formula (22.23) and the graph in fig. 11.7 it is seen that for $v \approx c$ there correspond to the values of the observation angle $\frac{1}{2}\pi < \vartheta < \pi - \sqrt{1 - v^2/c^2}$ values of the angle ϑ' close to zero. To angles ϑ' close to zero there correspond angles ϑ' close to π . This means that at velocities of motion of the body close to the velocity of light the radiation from the back of the body arrives at the photographic plate.

It should be stressed that this result refers to a body moving in a direction towards the photographing device ($\vartheta > \frac{1}{2}\pi$). In this case the intensity observed can be found from formula (22.17). Namely, we have

$$I_0(\vartheta) = I'_0(\vartheta') \frac{d\Omega'}{d\Omega} = I'_0(\vartheta') \frac{1 - \frac{v^2}{c^2}}{\left(1 + \frac{v}{c} \cos \vartheta\right)^2}.$$

If the angle $\vartheta < \pi - \sqrt{1 - v^2/c^2}$, then the observed intensity of light is small. This result has a simple meaning: for $v \sim c$ in the reference frame K the radiation is concentrated in a narrow cone with an aperture $\Delta\theta \sim \sqrt{1 - v^2/c^2}$ in the direction of the velocity of the body.

In general it can be said that if the object to be photographed moves with a velocity v and is observed at a certain angle ϑ , then one obtains the same picture as in photographing the object at rest but seen at an angle ϑ' . Thus, the image of a moving object on a photographic plate turns out to be the same as on a plate which is at rest with respect to the object. However, the object turns out to have undergone a rotation through an angle $\vartheta' - \vartheta$. Thus, in photographing a sphere one must obtain a circle on a photographic plate, in photographing a cube one must obtain a cube turned at a different angle, and so on.

Fig. 11.8 shows schematically how the image of a cube on a photographic plate must change as the relative velocity approaches the velocity of light. To the points A, B, C, D of the cube there correspond the points A', B', C', D' on the photographic plate. We see how the image progressively "rotates".

These considerations make particularly clear the lack of meaning of the terminology "an apparent contraction" in relation to the Lorentz contraction.

It should be noted that the entire picture becomes more complex if the object observed subtends a large solid angle. In this case to each point of the

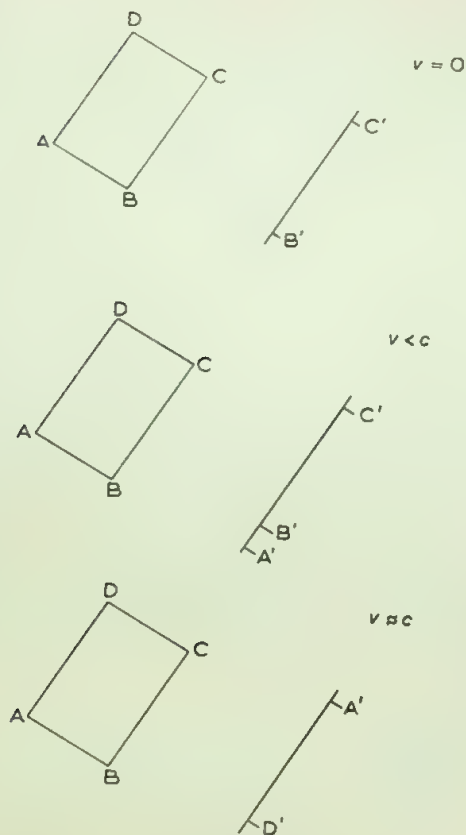


Fig. II.8

object there corresponds a proper observation angle ϑ and, consequently, also a proper rotation. However, in this case also the image of a sphere on a photographic plate must be a circle.

§ 23. The Lorentz force; the Lagrangian and the Hamiltonian for a particle moving in an electromagnetic field

In electrodynamics an expression was given for the force acting on a charge moving in an electromagnetic field. This force, called the Lorentz force, was deduced from experimental data. We shall now present a simple derivation of

the formula for the Lorentz force, based on the transformations of the electromagnetic field vectors.

Consider a charge e , moving with an arbitrary velocity \mathbf{v} with respect to the reference frame K . In the reference frame moving with the charge let there be an electric field \mathbf{E}' . Then the charge in the reference frame K' is acted upon by the force

$$\mathcal{F}' = e\mathbf{E}' \quad (23.1)$$

Our problem is to find the force acting on the charge in the reference frame K (in which it moves with the velocity \mathbf{v}). For this it is necessary to express the field \mathbf{E}' in terms of the electromagnetic field \mathbf{E} , and the force \mathcal{F}' in terms of the force \mathcal{F} in the reference frame K . The first transformation is given directly by the formulae of §21.

The formulae for the transformation of the force can be obtained in the following way. We find, first of all, the formulae for the transformation of the components of the Minkowski force (§12). The Minkowski force is a 4-vector, and its components transform according to the general formulae (11.1) – (11.4). In our case, in the reference frame moving with the charge the velocity $\mathbf{v}' = 0$ and, hence, $F'_4 = 0$. In this case

$$F_1 = \frac{F'_1}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad F_2 = F'_2, \quad F_3 = F'_3.$$

The components of the Minkowski force are connected with the components of the ordinary force \mathcal{F} by relations of the type (12.5).

In the reference frame moving with the charge $\mathbf{v}' = 0$, and

$$F'_1 = \mathcal{F}'_x, \quad F'_2 = \mathcal{F}'_y, \quad F'_3 = \mathcal{F}'_z.$$

In the reference frame K , by definition from §12,

$$F_1 = \frac{\mathcal{F}_x}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad F_2 = \frac{\mathcal{F}_y}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad F_3 = \frac{\mathcal{F}_z}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Hence we find easily *

$$\mathcal{F}_x = \mathcal{F}'_x, \quad \mathcal{F}_y = \mathcal{F}'_y \sqrt{1 - \frac{v^2}{c^2}}, \quad \mathcal{F}_z = \mathcal{F}'_z \sqrt{1 - \frac{v^2}{c^2}}. \quad (23.2)$$

Combining (23.1), (23.2) and the transformation formulae (21.7)–(21.8), we find the expression for the components of the force:

$$\mathcal{F}_x = \mathcal{F}'_x = eE'_x = eE_x, \quad (23.3)$$

$$\mathcal{F}_y = \mathcal{F}'_y \sqrt{1 - \frac{v^2}{c^2}} = eE'_y \sqrt{1 - \frac{v^2}{c^2}} = e \left(E_y - \frac{vH_z}{c} \right), \quad (23.4)$$

$$\mathcal{F}_z = \mathcal{F}'_z \sqrt{1 - \frac{v^2}{c^2}} = eE'_z \sqrt{1 - \frac{v^2}{c^2}} = e \left(E_z + \frac{vH_y}{c} \right). \quad (23.5)$$

In vector form

$$\mathcal{F} = e \left\{ \mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right\}. \quad (23.6)$$

Thus, the expression for the Lorentz force is obtained in a purely mathematical way from the general relations of the theory of relativity.

Let us write the equations of motion (12.3) for the case of the motion of a charged particle in an external electromagnetic field.

The first three equations are of the form

$$\frac{d\mathbf{p}}{dt} = e \left\{ \mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right\}. \quad (23.7)$$

In finding the fourth equation it must be taken into account that the work done by the force in the magnetic field is equal to zero (since $\mathbf{v} \cdot (\mathbf{v} \times \mathbf{H}) = \mathbf{H} \cdot (\mathbf{v} \times \mathbf{v}) = 0$) and, consequently

$$\frac{dE}{dt} = \mathcal{F} \cdot \mathbf{v} = e\mathbf{E} \cdot \mathbf{v}. \quad (23.8)$$

In the following paragraphs we shall consider some particular cases of the

* It should be stressed that the formulae presented are valid only for the transition from the reference frame K' moving together with the particle to the reference frame K .

motion of a particle in an electric and magnetic field. For what follows we shall need an expression for the Lorentz force in terms of electromagnetic potentials.

We have

$$\mathcal{F} = e \left\{ -\nabla\varphi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} + \frac{1}{c} \mathbf{v} \times (\nabla \times \mathbf{A}) \right\}$$

We make use of the formula (I.47) of vector analysis:

$$\begin{aligned} \nabla(\mathbf{A} \cdot \mathbf{v}) &= (\mathbf{A} \cdot \nabla)\mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{A} + \mathbf{v} \times (\nabla \times \mathbf{A}) + \mathbf{A} \times (\nabla \times \mathbf{v}) = \\ &= (\mathbf{v} \cdot \nabla)\mathbf{A} + \mathbf{v} \times (\nabla \times \mathbf{A}). \end{aligned}$$

Here we have taken into account that the differentiation with respect to coordinates is carried out at a constant value of the velocity \mathbf{v} .

Making use of this formula, we rewrite \mathcal{F} in the form

$$\begin{aligned} \mathcal{F} &= e \left\{ -\nabla\varphi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} + \frac{1}{c} \nabla(\mathbf{A} \cdot \mathbf{v}) - \frac{1}{c} (\mathbf{v} \cdot \nabla)\mathbf{A} \right\} = \\ &= e \nabla \left(\frac{\mathbf{A} \cdot \mathbf{v}}{c} - \varphi \right) - \frac{e}{c} \frac{d\mathbf{A}}{dt}, \end{aligned}$$

where the total derivative (see I.18) is

$$\frac{d\mathbf{A}}{dt} = \frac{\partial \mathbf{A}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{A}.$$

The equations of motion assume the form

$$\frac{d}{dt} \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}} = e \nabla \left\{ \frac{\mathbf{A} \cdot \mathbf{v}}{c} - \varphi \right\} - \frac{e}{c} \frac{d\mathbf{A}}{dt}. \quad (23.9)$$

These can be considered the Lagrange equations, if the Lagrangian is of the form

$$L = -mc^2 \sqrt{1 - \frac{v^2}{c^2}} - e\varphi + \frac{e}{c} \mathbf{A} \cdot \mathbf{v}. \quad (23.10)$$

Indeed, in this case the generalized momentum is

$$\mathbf{P} = \frac{\partial L}{\partial \mathbf{v}} = \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}} + \frac{e}{c} \mathbf{A} = \mathbf{p} + \frac{e}{c} \mathbf{A}. \quad (23.11)$$

Correspondingly the generalized force is

$$\mathbf{Q} = \frac{\partial L}{\partial \mathbf{r}} = \frac{e}{c} \nabla (\mathbf{A} \cdot \mathbf{v}) - e \nabla \varphi.$$

The Lagrange equation reads:

$$\frac{d}{dt} \frac{\partial L}{\partial \mathbf{v}} = \frac{\partial L}{\partial \mathbf{r}},$$

or

$$\frac{d}{dt} \mathbf{P} = \mathbf{Q}. \quad (23.12)$$

The substitution of the value \mathbf{P} and \mathbf{Q} into (23.12) again leads us to (23.9).

In the non-relativistic approximation the Lagrangian assumes the form

$$L \approx -mc^2 \left(1 - \frac{v^2}{2c^2} \right) + \frac{e}{c} \mathbf{A} \cdot \mathbf{v} - e\varphi = \frac{mv^2}{2} - e\varphi + \frac{e}{c} \mathbf{A} \cdot \mathbf{v}. \quad (23.13)$$

Here we have dropped the constant $(-mc^2)$, since Lagrange's equations include only the derivatives of L , and L itself is significant only with regard to its total derivative with respect to time. *

Comparing the Lagrangian of a particle in an electromagnetic field with the expression for the Lagrangian in an ordinary field of force,

$$L = \frac{mv^2}{2} - U,$$

we see that for the electromagnetic field the Lagrangian contains one more term, depending on the velocity and vector potential. Hence even in the non-relativistic approximation the Lagrangian of the electromagnetic field cannot

* L.D.Landau and E.M.Lifshitz, *Mechanics* (Pergamon, Oxford, 1960).

be written in the form of the difference between the kinetic and potential energy.

Let us find the Hamiltonian of a charged particle in an electromagnetic field. Obviously, we have

$$\begin{aligned}
 H = \sum_{i=x,y,z} \dot{q}_i P_i - L = \mathbf{v} \cdot \mathbf{P} - L = \mathbf{v} \cdot \left(\frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}} + \frac{e}{c} \mathbf{A} \right) + \\
 + mc^2 \sqrt{1 - \frac{v^2}{c^2}} + e\varphi - \frac{e}{c} \mathbf{A} \cdot \mathbf{v} = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} + e\varphi. \quad (23.14)
 \end{aligned}$$

In order to obtain the Hamiltonian the velocity \mathbf{v} should be expressed in terms of the generalized momentum \mathbf{P} . This is done most simply by means of the equations

$$(H - e\varphi)^2 = \frac{m^2 c^4}{1 - \frac{v^2}{c^2}}, \quad (23.15)$$

$$\frac{m^2 v^2}{1 - \frac{v^2}{c^2}} = \left(\mathbf{P} - \frac{e}{c} \mathbf{A} \right)^2, \quad (23.16)$$

which are obtained by raising (23.14) and (23.11) to the second power. Rewriting (23.16) in the form

$$\frac{m^2 v^2}{1 - \frac{v^2}{c^2}} + m^2 c^2 - m^2 c^2 = \frac{m^2 c^2}{1 - \frac{v^2}{c^2}} - m^2 c^2 = \left(\mathbf{P} - \frac{e}{c} \mathbf{A} \right)^2,$$

and comparing with (23.15), we obtain

$$(H - e\varphi)^2 = m^2 c^4 + \left(\mathbf{P} - \frac{e}{c} \mathbf{A} \right)^2 c^2,$$

or

$$H = \sqrt{m^2 c^4 + \left(\mathbf{P} - \frac{e}{c} \mathbf{A} \right)^2 c^2} + e\varphi. \quad (23.17)$$

As was to be expected from general considerations, the Hamiltonian in an electromagnetic field is, essentially, the same as the Hamiltonian in an electrostatic field: the magnetic field does not change the energy of the particle. In the non-relativistic approximation we obtain from (23.17)

$$H = mc^2 \sqrt{1 + \frac{\left(\mathbf{P} - \frac{e}{c} \mathbf{A}\right)^2}{m^2 c^2}} + e\varphi \approx mc^2 + \frac{\left(\mathbf{P} - \frac{e}{c} \mathbf{A}\right)^2}{2m} + e\varphi. \quad (23.18)$$

The last expression, if the rest energy is not taken into account, is the same as the classical expression for the Hamiltonian of a particle in an electromagnetic field (see (41.4) of Part I).

§ 24. The motion of particles in constant electric and magnetic fields

The simplest case of the motion of relativistic particles is their motion in constant electric and magnetic fields. At the same time, the motion of charged particles in electric and magnetic fields is of very great practical interest. It is sufficient to quote some examples: the investigation of the motion of electrons in such fields made it possible to test to a high degree of accuracy the relativistic expression for momentum; the relativistic formulae determining the law of motion of particles in electric and magnetic fields represent the basis for the design of contemporary nuclear particle accelerators, allowing one to obtain particles of relativistic energies. The investigation of the motion of very fast particles in a Wilson cloud chamber placed in a magnetic field allows the determination of their momentum. The combination of the measurements of the momentum and the energy (carried out, for example, on the basis of the ionization produced by the particles) makes it possible to determine the mass of the particles.

Consider, first of all, the motion of a relativistic particle in an electric field \mathbf{E} which is constant in time and uniform in space. We choose the direction of the vector \mathbf{E} to be the x -axis. According to § 12, the equations of motion are of the form

$$\frac{dp_x}{dt} = eE_x,$$

$$\frac{dp_y}{dt} = \frac{dp_z}{dt} = 0,$$

$$\frac{dE}{dt} = eE_x v_x = eE_x \frac{dx}{dt}.$$

As the first example we shall consider the motion of a charge in a transverse electric field. At the time $t = 0$ let the charge be at the point $x = 0$ and have a momentum $p_x = 0, p_y = p_0, p_z = 0$ and energy

$$E_0 = \sqrt{p_0^2 c^2 + m^2 c^4}.$$

This means that at the initial instant the charge was moving in a direction perpendicular to the field. The integration of the equations for the components of the momentum in the constant electric field is carried out directly and, taking into account the initial conditions, gives

$$p_x = eE_x t, \quad p_y = p_0, \quad p_z = 0. \quad (24.1)$$

The equation for the energy is also integrated directly:

$$E = eE_x x + E_0.$$

On the other hand,

$$\begin{aligned} E &= \sqrt{(p_x^2 + p_y^2 + p_z^2)c^2 + m^2 c^4} = \\ &= \sqrt{[(eE_x t)^2 + p_0^2]c^2 + m^2 c^4} = \sqrt{(ecE_x t)^2 + E_0^2}. \end{aligned}$$

Comparing the two expressions for the energy, we can write

$$\sqrt{(ecE_x t)^2 + E_0^2} = eE_x x + E_0. \quad (24.2)$$

Formulae (24.1) show that the motion is planar and takes place in the (xy) -plane. In order to find the trajectory one can write the relation

$$\frac{p_x}{p_y} = \frac{mv_x}{\sqrt{1-v^2/c^2}} \frac{\sqrt{1-v^2/c^2}}{mv_y} = \frac{v_x}{v_y} = \frac{dx}{dy}. \quad (24.3)$$

Substituting the value of p_x/p_y from (24.1) into (24.3), we obtain

$$\frac{dx}{dy} = \frac{eE_x t}{p_0}. \quad (24.4)$$

To eliminate the time we make use of formula (24.2), which gives

$$t = \frac{\sqrt{(eE_x x + E_0)^2 - E_0^2}}{ecE_x}.$$

Substituting this expression into (24.4), we arrive at the differential equation for the trajectory:

$$\frac{dx}{dy} = \frac{\sqrt{(eE_x x + E_0)^2 - E_0^2}}{cp_0}.$$

Integration gives the equation of the trajectory

$$\frac{y}{cp_0} = \int \frac{dx}{\sqrt{(eE_x x + E_0)^2 - E_0^2}} = \frac{1}{eE_x} \operatorname{arccosh} \frac{eE_x x}{E_0},$$

or

$$x = \frac{E_0}{eE_x} \cosh \frac{eE_x y}{cp_0}. \quad (24.5)$$

Eq. (24.5) shows that in the transverse electric field the charge moves on a catenary. For $v \ll c$ we can write

$$E_0 \approx mc^2, \quad p_0 \approx mv_0 \quad \text{and} \quad x \approx \frac{mc^2}{eE_x} \cosh \frac{eE_x y}{mv_0 c}. \quad (24.6)$$

If the argument of the hyperbolic cosine, containing c in the denominator, is small, then, expanding (24.6) into a series we arrive at the classical formula for the trajectory, which we have already found earlier (see § 39 of Part I).

The second important case is the motion of a charge in the direction of the electric field, i.e. the case of the initial conditions

$$p_x^{(0)} = 0, \quad p_y^{(0)} = 0, \quad p_z^{(0)} = 0, \quad E^{(0)} = mc^2.$$

In this case the field will be considered to be constant in time, but varying arbitrarily in space along the x -axis.

The integration of the equations for the components of the momentum gives

$$p_x = eE_x(x)t, \quad p_y = 0, \quad p_z = 0. \quad (24.7)$$

We write the equation for the energy in the form

$$\frac{dE}{dt} = -e \frac{d\varphi}{dx} \frac{dx}{dt} = -e \frac{d\varphi}{dt},$$

where φ is the electrostatic potential. Whence we find the integral of the energy:

$$E + e\varphi = \text{const} = mc^2 + e\varphi_0,$$

or

$$\Delta E = mc^2 \left[\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} - 1 \right] = eV, \quad (24.8)$$

where V is the difference of potential through which the particle has passed, and ΔE is the corresponding change in the energy. The velocity of the particle which has passed through the difference of potential V is equal to

$$v = \sqrt{\frac{2eV}{m} \frac{1 + \frac{e}{m} \frac{V}{2c^2}}{\left(1 + \frac{e}{m} \frac{V}{c^2}\right)^2}}. \quad (24.9)$$

This expression is used in practice for the calculation of the velocity of a particle accelerated by an electric field (fig. II.9).

For $eV/mc^2 \ll 1$ formula (24.9) reduces to the classical expression (39.1) of Part I. On the contrary, for $eV/mc^2 \gg 1$ the velocity of the particle tends to a constant limit $v \rightarrow c$ as the potential increases.

If the field is uniform in space, then the dependence of the velocity and coordinate on time is obtained directly from (24.7). Namely, from (24.7) we have

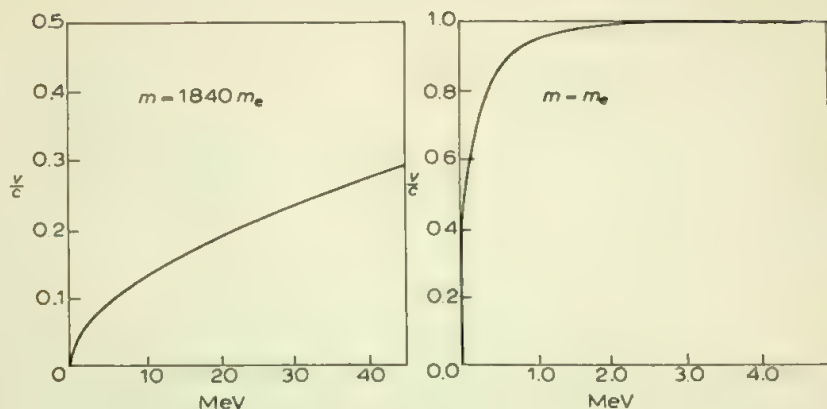


Fig. 11.9

$$v = \frac{c \left(\frac{eE_x t}{mc} \right)}{\sqrt{1 + \left(\frac{eE_x t}{mc} \right)^2}}. \quad (24.10)$$

Formula (24.10) gives the law of uniformly-accelerated motion in relativistic mechanics. A uniformly accelerated motion in the theory of relativity is understood to be the following: Let us introduce a number of inertial reference frames K'_1, K'_2, K'_3, \dots , moving with velocities equal to those of the motion of the particle at different instants. Each of the reference frames K' is called a frame instantaneously accompanying the particle. In such a reference frame accompanying the particle its velocity is equal to zero for an instant.

Hence, according to (11.12) and (11.16), the components of the 4-acceleration in K' are equal to

$$w'_x = \dot{v}'_x = \dot{v} = \frac{eE_x}{m}, \quad w'_\tau = 0.$$

By virtue of (11.1) and (11.13), we have in a reference frame at rest

$$w_x = \frac{w'_x}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{eE_x}{m \sqrt{1 - \frac{v^2}{c^2}}} = \frac{\dot{v}}{1 - \frac{v^2}{c^2}} + \frac{v^2 \dot{v}}{c^2 \left(1 - \frac{v^2}{c^2} \right)^2}.$$

Whence we find the quantity \dot{v} , which we shall call the acceleration:

$$\dot{v} = \frac{eE_x}{m} \left(1 - \frac{v^2}{c^2}\right)^{3/2}. \quad (24.11)$$

Integrating (24.11) we again arrive at formula (24.10). Integrating (24.10) we obtain the law for motion

$$x - x_0 = \frac{mc^2}{eE_x} \sqrt{1 + \left(\frac{eE_x}{mc} t\right)^2}. \quad (24.12)$$

This is the equation of a hyperbola. Hence in relativistic mechanics the motion in a constant field is often called hyperbolic, as distinct from the parabolic trajectory of classical mechanics.

Finally, consider the motion of a particle in a constant and uniform magnetic field \mathbf{H} . We choose the direction of the latter to be the z -axis. Then the equations of motion take the form

$$\frac{dp_x}{dt} = \frac{e}{c} v_y H, \quad \frac{dp_y}{dt} = -\frac{e}{c} v_x H, \quad \frac{dp_z}{dt} = 0, \quad \frac{dE}{dt} = 0. \quad (24.13)$$

Integrating the equation for energy, we have

$$E = \text{const} = E_0.$$

By means of (13.2) we obtain

$$\frac{dp_x}{dt} = \frac{d}{dt} \left(\frac{E}{c^2} v_x \right) = \frac{E_0}{c^2} \frac{dv_x}{dt}, \quad \frac{dp_y}{dt} = \frac{E_0}{c^2} \frac{dv_y}{dt}.$$

Hence the equations for the x and y components of the momentum can be written in the form

$$\begin{aligned} \frac{1}{v_y} \frac{dv_x}{dt} &= \frac{ecH}{E_0}, \\ \frac{1}{v_x} \frac{dv_y}{dt} &= -\frac{ecH}{E_0}. \end{aligned} \quad (24.14)$$

Eqs. (24.14) can be satisfied by setting

$$v_x = a \cos(\omega t + \alpha), \quad v_y = -a \sin(\omega t + \alpha). \quad (24.15)$$

For ω we find

$$\omega = \frac{ecH}{E_0}. \quad (24.16)$$

From (24.15)

$$v_x^2 + v_y^2 = a^2 = (v_{\perp})^2 = (v_{\perp})_0^2 = \text{const.}$$

The quantity $(v_{\perp})_0$, which represents the initial velocity of motion of a particle in the (xy) -plane, remains constant in time.

The second integration leads to the equation of the trajectory in the (xy) -plane

$$x = x_0 + \frac{(v_{\perp})_0}{\omega} \sin(\omega t + \alpha), \quad y = y_0 + \frac{(v_{\perp})_0}{\omega} \cos(\omega t + \alpha).$$

The particle moves with a constant velocity along the z -axis, as is seen from (24.13):

$$z = z_0 + (v_z)_0 t.$$

If, in particular, at the initial instant the charge had no velocity along the z -axis, i.e. $(v_z)_0 = 0$, then its trajectory represents a circle in the plane (xy) the radius of which is equal to

$$R = \frac{(v_{\perp})_0}{\omega} = \frac{(v_{\perp})_0 E_0}{ecH} = \frac{cp_0}{eH}, \quad (24.17)$$

where p_0 is the initial momentum.

The frequency ω of rotation in the circle is proportional to the strength of the magnetic field and depends on the energy E_0 which remains constant during the motion. At small energies

$$E_0 \approx mc^2 \quad \text{and} \quad \omega \approx eH/mc, \quad (24.18)$$

i.e. ω reduces to the cyclotron frequency determined in §39 of Part I, which does not depend on the energy of the particle.

Formulae (24.16) and (24.17) are of very great importance in contemporary nuclear physics and technology.

The measurement of the deflection of particles in a magnetic field allows one to find their momentum by means of formula (24.17).

Formula (24.16) is the basis for the design of contemporary cyclotrons, which allow one to obtain heavy particles (protons, deuterons and α -particles) with relativistic velocities.

As is well known, in the cyclotron particles move in circles in a magnetic field between two dees to which an alternating voltage is applied. Particles pass through the gap between the dees in an accelerating electric field. Thus, after describing a semicircle with a constant velocity the particle is accelerated and describes the next semicircle with a new value of the velocity, and so on. For continuous acceleration the field in the accelerating gap must be in a definite phase at the instant when the particle enters it. So long as the particles do not acquire relativistic velocities the frequency of the electric field applied to the dees is determined by formula (24.18) and does not depend on the energy of the particle. If, however, the velocity of the particles attains relativistic values, then, according to (24.16), the frequency of their rotation turns out to vary with energy. Hence, for a further acceleration of the particles, it is necessary to change the frequency of the accelerating field in correspondence with formula (24.16). Cyclotrons operating in this varying frequency mode are called relativistic cyclotrons or synchrotrons.

The design of synchrotrons was first suggested in 1944 by V.I. Veksler, who showed that owing to the particular properties of the motion of charged particles (the so-called phase stability), the synchrotron accelerates particles arriving at the accelerator chamber with different initial phases of motion.

The acceleration of light particles — e.g. electrons — takes place in the relativistic mode at relatively small energies.

One of the most important types of accelerators is the induction accelerator or the betatron. In the betatron electrons move in a magnetic field with axial symmetry between the poles of an electromagnet. If the magnetic field were constant in time, then the motion of an electron with constant velocity would take place on a circle of radius R given by formula (24.17). In the betatron, however, the strength of the magnetic field varies in time. For concreteness we shall assume that the strength of the magnetic field increases in time. This variation of the strength of the magnetic field entails:

- 1) The appearance of an induced electric field with a strength E determined by the formula

$$\oint \mathbf{E} \, d\mathbf{l} = -\frac{1}{c} \frac{\partial}{\partial t} \int \mathbf{H} \cdot d\mathbf{S},$$

or

$$E_{\varphi} = -\frac{1}{2\pi R c} \frac{d\Phi}{dt},$$

where $\Phi = \int \mathbf{H} \cdot d\mathbf{S}$ is the flux of induction through the area of a circular orbit of radius R . It is clear that, because of the axial symmetry of the magnetic field, the electric field is directed along a tangent to the circle of radius R . Consequently, the electron will be acted upon by a force $(-eE_{\varphi})$, which is also directed along a tangent to this circle. According to (24.17), the increase in the momentum of the electron will correspond to an increase in the radius of the orbit, i.e. to a tendency of the electron to move outwards on a spiral.

2) A decrease in the radius of the circle in (24.17) as the strength \mathbf{H} increases. This corresponds to a tendency of the electron to move inwards on a spiral.

If the rate of change of \mathbf{H} in time (and, as will be seen from what follows, in space) is chosen in such a way that the two tendencies exactly compensate each other, the electron will move on a circle of constant radius R with increasing momentum. This is the so-called betatron mode of the motion of a particle.

We now consider the conditions under which such a mode will occur.

The change of the momentum of the electron can be written in the form

$$\frac{dp}{dt} = -eE_{\varphi} = \frac{e}{2\pi c R} \frac{d\Phi}{dt}. \quad (24.19)$$

If it is assumed that the electron moves on a circle of constant radius, then on integrating (24.19) we obtain

$$p = \frac{e}{2\pi c R} \Phi.$$

The integration constant is set equal to zero under the assumption that the conditions $H = 0$ and $v = 0$ hold at the instant $t = 0$.

If p is substituted into (24.17), then the condition of constancy of the radius of the orbit in time can be written in the form

$$R = \frac{cp}{eH} = \frac{1}{2\pi} \frac{\Phi}{RH}. \quad (24.20)$$

The flux of magnetic induction through the area of the orbit is equal to

$$\Phi = \pi R^2 \bar{H},$$

where \bar{H} is the mean field inside the orbit.

Then the condition (24.20) reduces to the equality

$$H_{r=R} = \frac{1}{2} \bar{H}.$$

This indicates that for the motion of an accelerated electron on a circular orbit it is necessary to produce a magnetic field not only varying in time but also non-uniform in space. The field on the orbit must be equal to one half of the mean strength inside the orbit. For this requirement to be fulfilled the field must decrease with increasing radius r . It is clear that the acceleration of particles in the betatron is of intermittent character — it takes place only when the magnetic field increases in time. We cannot consider here the problems of the stability of the motion of particles in a betatron orbit and the details of the design of actual accelerators. *

§25. A system of weakly interacting charged particles

We can now return to a system of particles in relativistic mechanics and consider the case mentioned in §15: that of particles connected by the electromagnetic interaction. **

We have already shown that a potential energy of interaction cannot be introduced for a system of interacting particles, since the retarded interaction depends not on the relative position of the particles at a given instant but on their motion during the preceding time. Moreover, in an accelerated motion charged particles emit radiation and a part of the energy leaves the system, so that the system as a whole is non-conservative.

However, it turns out that if the motion of the particles is sufficiently slow, so that $v \ll c$, then to a certain approximation one can introduce the

* See, for example: A.P.Grünberg, *Metody uskoreniya zaryazhonykh chastits (Methods of accelerating charged particles)* (Gostekhizdat, Moscow, 1960); M.S.Livingston and J.P.Blewett, *Particle accelerators* (McGraw-Hill, New York, 1962).

** V.A.Foek, *The theory of space, time and gravitation* (Pergamon, London, 1964); L.D.Landau and E.M.Lifshitz, *The classical theory of fields* (Pergamon, London, 1962).

notion of the interaction of a system of charges which depends only on their mutual separations. This allows one to characterize the state of a system of moving charges by means of mechanical quantities and to consider the motion of the system, according to the laws of mechanics without a direct connection with the state of the electromagnetic field.

Indeed, if the charges are moving with small velocities, so that the retardation can be completely disregarded, then their energy of interaction is expressed by the formula of electrostatics:

$$U = \sum_{i < k} \frac{e_i e_k}{r_{ik}}, \quad (25.1)$$

where r_{ik} is the fixed distance between the charges i and k . For an arbitrarily chosen k th charge the Lagrangian can be written in the form

$$L_k = \frac{1}{2} m_k v_k^2 - e_k \varphi_k, \quad (25.2)$$

where φ_k is the potential of the field acting on the charge k .

The Lagrangian of a system of charges is obtained by a simple summation:

$$L = \sum_k L_k = \sum_k \left(\frac{1}{2} m_k v_k^2 - e_k \varphi_k \right). \quad (25.3)$$

Knowing the Lagrangian, one can find the equations of motion of the system.

Further calculations will show also that in successive approximations of the expansion in powers of the ratio v/c , up to terms of the order of v^2/c^2 , the Lagrangian of the system can be found. This will also allow us to carry out the programme of a purely mechanical description of a system of charges previously mentioned.

At the same time it is clear that, if terms of the order of $(v/c)^3$ are not discarded, then such a description will, generally speaking, become impossible. Indeed, according to the results of § 27 of Part I, the dipole radiation of a system is determined by a quantity of the order of $1/c^3$. The retention of the higher terms of the expansion of potentials corresponds to taking account of the dipole radiation. This makes the mechanical approach to the treatment of the state of the system illegitimate.

In systems in which there is no dipole radiation the expansion of the potentials can be carried out up to terms of a low order of magnitude as follows. We write the Lagrangian of the k th particle taking into account the motion of the charges in the system (see eq. 23.10):

$$L_k = -m_k c^2 \sqrt{1 - \frac{v_k^2}{c^2}} - e_k \varphi_k + \frac{e_k}{c} (\mathbf{v}_k \cdot \mathbf{A}_k), \quad (25.4)$$

where φ_k and \mathbf{A}_k are the potentials of the field at that point where the k th charge is located. These potentials, taking into account the finiteness of the velocity of propagation of interaction, can be written in the form of retarded potentials (see §24 of Part I):

$$\varphi(\mathbf{r}_k, t) = \int \frac{\rho\left(\mathbf{r}', t - \frac{|\mathbf{r}_k - \mathbf{r}'|}{c}\right)}{|\mathbf{r}_k - \mathbf{r}'|} dV',$$

$$\mathbf{A}(\mathbf{r}_k, t) = \frac{1}{c} \int \frac{\mathbf{j}\left(\mathbf{r}', t - \frac{|\mathbf{r}_k - \mathbf{r}'|}{c}\right)}{|\mathbf{r}_k - \mathbf{r}'|} dV'.$$

For a slow motion ($v \ll c$) the charge density and current density can be expanded in a series of powers of the delay time. An important difference of this expansion from the analogous formulae of §26 of Part I lies in the fact that we are interested in potentials at a point \mathbf{r}_k located within the limits of the system of charges considered. Hence the total delay time cannot be resolved into the proper retardation and the retardation of the system, and the expansion must be carried out with respect to the total retardation:

$$\tau = \frac{|\mathbf{r}_k - \mathbf{r}|}{c}.$$

We then have

$$\begin{aligned} \rho\left(\mathbf{r}', t - \frac{|\mathbf{r}_k - \mathbf{r}'|}{c}\right) &\approx \rho(\mathbf{r}', t) - \frac{|\mathbf{r}_k - \mathbf{r}'|}{c} \frac{\partial}{\partial t} \rho(\mathbf{r}', t) + \\ &+ \frac{1}{2} \frac{|\mathbf{r}_k - \mathbf{r}'|^2}{c^2} \frac{\partial^2}{\partial t^2} \rho(\mathbf{r}', t), \\ \mathbf{j}\left(\mathbf{r}', t - \frac{|\mathbf{r}_k - \mathbf{r}'|}{c}\right) &= \mathbf{j}(\mathbf{r}', t) + O(1/c). \end{aligned}$$

Terms $O(1/c)$ are neglected because of the factors $1/c$ in $\mathbf{A}(\mathbf{r}_k, t)$ above and in eq. (25.4). Hence

$$\begin{aligned}
 \varphi(\mathbf{r}_k, t) &\approx \int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r}_k - \mathbf{r}'|} dV' - \frac{1}{c} \int \frac{\partial \rho(\mathbf{r}', t)}{\partial t} dV' + \\
 &\quad + \frac{1}{2c^2} \int |\mathbf{r}_k - \mathbf{r}'| \frac{\partial^2 \rho(\mathbf{r}', t)}{\partial t^2} dV' = \\
 &\int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r}_k - \mathbf{r}'|} dV' - \frac{1}{c} \frac{\partial}{\partial t} \int \rho(\mathbf{r}', t) dV' + \frac{1}{2c^2} \frac{\partial^2}{\partial t^2} \int |\mathbf{r}_k - \mathbf{r}'| \rho(\mathbf{r}', t) dV'.
 \end{aligned}$$

The order of the differentiation and integration can be changed, since the vector \mathbf{r}_k is fixed, and \mathbf{r}' is a set of three independent variables.

Since the integral in the second term is taken at a time t , it represents the total charge of the system. Correspondingly,

$$\frac{\partial}{\partial t} \int \rho(\mathbf{r}', t) dV' = \frac{\partial e}{\partial t} = 0.$$

Finally, we find

$$\varphi(\mathbf{r}_k, t) = \int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r}_k - \mathbf{r}'|} dV' + \frac{1}{2c^2} \frac{\partial^2}{\partial t^2} \int |\mathbf{r}_k - \mathbf{r}'| \rho(\mathbf{r}', t) dV' \quad (25.5)$$

$$\mathbf{A}(\mathbf{r}_k, t) = \frac{1}{c} \int \frac{\mathbf{j}(\mathbf{r}', t)}{|\mathbf{r}_k - \mathbf{r}'|} dV' \quad (25.6)$$

Further calculations become clearer if one considers point charges. Let

$$\rho(\mathbf{r}', t) = \sum_i' e_i \delta(\mathbf{r}' - \mathbf{r}_i(t)). \quad (25.7)$$

The prime on the summation sign indicates that in the sum the term $i = k$ is absent since we do not take into account the field due to the k th charge itself. In what follows we shall omit the prime on the sum so as not to encumber the formulae, but shall imply it in all summations over charges.

Substituting the expression (25.7) for $\rho(\mathbf{r}', t)$ into (25.5), we have

$$\begin{aligned}
\varphi(\mathbf{r}_k, t) &= \sum \varphi_i = \\
&= \sum \frac{e_i}{|\mathbf{r}_k - \mathbf{r}_i(t)|} + \frac{1}{2c^2} \frac{\partial^2}{\partial t^2} \sum e_i |\mathbf{r}_k - \mathbf{r}_i(t)| = \\
&= \sum \frac{e_i}{|\mathbf{r}_k - \mathbf{r}_i(t)|} + \frac{1}{2c^2} \sum e_i \frac{\partial^2}{\partial t^2} |\mathbf{r}_k - \mathbf{r}_i(t)|. \quad (25.8)
\end{aligned}$$

We stress that as a result of integration (elimination of the δ -function) the radius vector of the i th charge $\mathbf{r}_i(t)$, depending explicitly on time, appears instead of \mathbf{r}' in the corresponding expressions. Hence the result of integration with respect to the variables \mathbf{r}' must be differentiated with respect to time.

Analogously, assuming

$$\mathbf{j}(\mathbf{r}', t) = \sum_i' e_i \mathbf{v}_i \delta(\mathbf{r}' - \mathbf{r}_i(t)), \quad (25.9)$$

we have from (25.6)

$$\mathbf{A}(\mathbf{r}_k, t) = \sum \mathbf{A}_i = \sum \frac{e_i \mathbf{v}_i}{c |\mathbf{r}_k - \mathbf{r}_i|}. \quad (25.10)$$

Here φ_i and \mathbf{A}_i are the potentials produced at an instant t at a point \mathbf{r}_k by the i th charge.

The expression (25.8) for the scalar potential taking into account the retardation (the second term) contains the second derivative with respect to time of the vector \mathbf{r}_i , i.e. the acceleration of the i th particle producing a field at the point \mathbf{r}_k . However, only the coordinates and velocities of the particles can enter into the Lagrangian. Hence it is advisable to carry out the gauge transformation (see §11 of Part I) and to choose the function ψ in such a way that the second term will be absent in the scalar potential. That is to say, assuming

$$\begin{aligned}
\varphi_i &\rightarrow \varphi_i' + \frac{1}{c} \frac{\partial \psi_i}{\partial t}, \\
\mathbf{A}_i &\rightarrow \mathbf{A}_i' - \nabla \psi_i,
\end{aligned}$$

where, using (I.12'),

$$\begin{aligned}
 \psi_i &= \frac{e_i}{2c} \frac{\partial}{\partial t} |\mathbf{r}_k - \mathbf{r}_i(t)| = \frac{e_i}{2c} \nabla_{\mathbf{r}_i} |\mathbf{r}_k - \mathbf{r}_i| \cdot \frac{\partial \mathbf{r}_i}{\partial t} = \\
 &= \frac{e_i}{2c} \left\{ -\frac{(\mathbf{r}_k - \mathbf{r}_i)}{|\mathbf{r}_k - \mathbf{r}_i|} \right\} \frac{\partial \mathbf{r}_i}{\partial t} = -\frac{e_i}{2c} \frac{(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_i}{|\mathbf{r}_k - \mathbf{r}_i|}, \quad (25.11)
 \end{aligned}$$

we find from (25.5)

$$\varphi' = \sum \varphi'_i = \sum_i \frac{e_i}{|\mathbf{r}_k - \mathbf{r}_i|}. \quad (25.12)$$

In differentiating with respect to time, the position of the observation point \mathbf{r}_k is fixed, the gradient is directed from the charge i to the observation point, and \mathbf{v}_i denotes the velocity of motion of the i th charge at an instant t .

Correspondingly, for A' we obtain

$$A'_i = A_i + \nabla_{\mathbf{r}_k} \psi_i = \sum_i \frac{e_i \mathbf{v}_i}{c |\mathbf{r}_k - \mathbf{r}_i|} - \nabla \frac{e_i}{2c} \frac{(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_i}{|\mathbf{r}_k - \mathbf{r}_i|} \quad (25.13)$$

According to the formula

$$\nabla_{\mathbf{r}_k} \frac{(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{a}}{|\mathbf{r}_k - \mathbf{r}_i|} = \frac{\mathbf{a}}{|\mathbf{r}_k - \mathbf{r}_i|} - \frac{(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{a}}{|\mathbf{r}_k - \mathbf{r}_i|^3} (\mathbf{r}_k - \mathbf{r}_i),$$

where \mathbf{a} is a constant vector, we find

$$\nabla_{\mathbf{r}_k} \frac{(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_i}{|\mathbf{r}_k - \mathbf{r}_i|} = \frac{\mathbf{v}_i}{|\mathbf{r}_k - \mathbf{r}_i|} - \frac{(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_i}{|\mathbf{r}_k - \mathbf{r}_i|^3} (\mathbf{r}_k - \mathbf{r}_i).$$

Hence for the total vector potential we obtain

$$A'(\mathbf{r}_k, t) = \sum A'_i = \sum_i \left(\frac{e_i \mathbf{v}_i}{2c |\mathbf{r}_k - \mathbf{r}_i|} + \frac{e_i}{2c} \frac{(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_i}{|\mathbf{r}_k - \mathbf{r}_i|^3} (\mathbf{r}_k - \mathbf{r}_i) \right). \quad (25.14)$$

The expression (25.12) and (25.14) which have been found for φ' and A' must be substituted into the Lagrangian (25.4) of the k th particle. One must first expand the first term in it in a series in powers of v^2/c^2 and retain terms of the same low order of magnitude as the terms retained in φ' and A' . This gives

$$m_k c^2 \sqrt{1 - \frac{v_k^2}{c^2}} \approx m_k c^2 \left(1 - \frac{v_k^2}{2c^2} - \frac{1}{8} \frac{v_k^4}{c^4} \right). \quad (25.15)$$

As a result we find

$$\begin{aligned} L_k &= -m_k c^2 + \frac{m_k v_k^2}{2} + \frac{1}{8} \frac{m_k v_k^4}{c^2} - e_k \varphi' + \frac{e_k (\mathbf{v}_k \cdot \mathbf{A}')}{c} = \\ &= -m_k c^2 + \frac{m_k v_k^2}{2} + \frac{1}{8} \frac{m_k v_k^4}{c^2} - \sum_i \left\{ \frac{e_i e_k}{|\mathbf{r}_k - \mathbf{r}_i|} - \right. \\ &\quad \left. - \frac{1}{2c^2} \frac{e_i e_k \mathbf{v}_i \cdot \mathbf{v}_k}{|\mathbf{r}_k - \mathbf{r}_i|} - \frac{e_i e_k [(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_i] [(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_k]}{2c^2 |\mathbf{r}_k - \mathbf{r}_i|^3} \right\}. \end{aligned} \quad (25.16)$$

The first term of the Lagrangian L_k refers to the particle at rest, the second term has the meaning of the kinetic energy in the approximation of classical mechanics, and the third term has the meaning of the relativistic correction to the kinetic energy. The term in the brackets depends only on the instantaneous positions and velocities of the particles.

The quantity

$$U_{ik} = \frac{e_i e_k}{|\mathbf{r}_k - \mathbf{r}_i|} - \frac{1}{2c^2} \frac{e_i e_k \mathbf{v}_i \cdot \mathbf{v}_k}{|\mathbf{r}_k - \mathbf{r}_i|} - \frac{e_i e_k [(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_i] [(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_k]}{2c^2 |\mathbf{r}_k - \mathbf{r}_i|^3}, \quad (25.17)$$

depending on the distance between the i th and k th particle at a given instant can be considered as the generalized energy of interaction between the particles. The first term has an obvious meaning: it is the potential energy of interaction between two charges, the i th one and the k th one, at rest. The other two terms, proportional to $\mathbf{v}_i \cdot \mathbf{v}_k / c^2$, represent the correction to the energy of interaction taking into account the motion of the charges and the retardation. However, it is clear that, although U_{ik} is the energy of interaction, it does not have the meaning of a potential energy depending only on the position of the particles.

The expression (25.17) is completely symmetrical in the two charges. Hence it is easy to write the Lagrangian for a system of particles. Namely,

$$L = \sum_k \left\{ -m_k c^2 + \frac{m_k v_k^2}{2} + \frac{m_k v_k^4}{8c^2} \right\} - \sum_{k>i} \left\{ \frac{e_i e_k}{|\mathbf{r}_k - \mathbf{r}_i|} - \frac{e_i e_k}{2c^2} \frac{\mathbf{v}_i \cdot \mathbf{v}_k}{|\mathbf{r}_k - \mathbf{r}_i|} - \frac{e_i e_k}{2c^2} \frac{[(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_i][(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_k]}{|\mathbf{r}_k - \mathbf{r}_i|^3} \right\} = L_1 + L_2, \quad (25.18)$$

where L_1 is the Lagrangian of a system of charges when relativistic corrections and the retardation are disregarded, given by formula (25.3), and L_2 is an addition to it, found with an accuracy up to terms of order less than $(v/c)^3$:

$$L_2 = \frac{1}{4c^2} \left\{ \sum_k \frac{m_k v_k^4}{2} + 2 \sum_{k>i} e_i e_k \left[\frac{\mathbf{v}_i \cdot \mathbf{v}_k}{|\mathbf{r}_k - \mathbf{r}_i|} + \frac{[(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_i][(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_k]}{|\mathbf{r}_k - \mathbf{r}_i|^3} \right] \right\} \quad (25.19)$$

Knowing the Lagrangian, one can find the energy, mass and momentum of the system.

The energy of the system is found according to the usual rules and is equal to

$$E_{\text{tot}} = \sum_k \mathbf{v}_k \cdot \frac{\partial L}{\partial \mathbf{v}_k} - L = E_0 + E_1 + E_2, \quad (25.20)$$

where

$$E_0 = \sum_k m_k c^2, \quad E_1 = \frac{1}{2} \sum_k m_k v_k^2 + \sum_{k>i} \frac{e_i e_k}{|\mathbf{r}_k - \mathbf{r}_i|},$$

and E_2 is the relativistic correction to the energy, which is equal to

$$E_2 = \frac{3}{8} \sum_k \frac{m_k v_k^4}{c^2} + \frac{1}{2c^2} \left\{ \sum_{k>i} e_i e_k \left[\frac{\mathbf{v}_i \cdot \mathbf{v}_k}{|\mathbf{r}_k - \mathbf{r}_i|} + \frac{[(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_i][(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_k]}{|\mathbf{r}_k - \mathbf{r}_i|^3} \right] \right\} \quad (25.21)$$

We see, first of all, that the energy of the system cannot be written in the form of a sum of the kinetic and potential energies. The relativistic correc-

tion to the energy E_2 depends both on the coordinates and on the velocities, so that taking account of this correction a potential energy for the system does not exist. Only for charges at rest or, more precisely, charges moving so slowly that quantities of the order of v^2/c^2 can completely be disregarded, can the quantity E_2 be discarded and use be made of the potential energy (25.1).

With the usual definition of mass, we find the mass of the system

$$M = \frac{E_{\text{tot}}}{c^2} = \sum_k m_k + \frac{E_1 + E_2}{c^2}. \quad (25.22)$$

Thus, the mass of the system is made up of the rest masses of the particles and the masses due to the kinetic energy of interaction of the particles of the system (in the approximation of charges at rest — the potential energy). The mass M obviously possesses no additive properties and is not equal to the sum of the masses of individual particles. The energy E of the system and its mass M are conserved. However, a conservation law cannot be written for individual terms entering into E_{tot} and M .

The momentum \mathbf{P} of the system is by definition equal to

$$\mathbf{P} = \sum_k \frac{\partial L_k}{\partial \mathbf{v}_k} = \mathbf{P}_1 + \mathbf{P}_2, \quad (25.23)$$

where $\mathbf{P}_1 = \sum_k m_k \mathbf{v}_k$ is the usual value of the momentum in classical mechanics and \mathbf{P}_2 is the relativistic correction to it:

$$\begin{aligned} \mathbf{P}_2 = & \frac{1}{2c^2} \sum m_k v_k^2 \mathbf{v}_k + \frac{1}{2c^2} \sum_{k>i} e_i e_k \left[\frac{\mathbf{v}_i}{|\mathbf{r}_k - \mathbf{r}_i|} + \right. \\ & \left. + \frac{(\mathbf{r}_k - \mathbf{r}_i) \cdot \mathbf{v}_i}{|\mathbf{r}_k - \mathbf{r}_i|^3} (\mathbf{r}_k - \mathbf{r}_i) \right]. \end{aligned} \quad (25.24)$$

We see that the correction \mathbf{P}_2 to the momentum depends on the coordinates of the particles of the system.

It is easy to show that in the approximation considered one can introduce the notion of the centre of mass of the system, which does not exist in an arbitrary system of interacting particles. From the definition of the velocity of the centre of mass (15.16) $\mathbf{V}_{\text{c.m.}} = c^2 \mathbf{P} / E_{\text{tot}}$ it can be seen that the vector $\mathbf{V}_{\text{c.m.}}$ can be written in the form of the derivative of the radius vector of the

centre of mass with respect to time:

$$\mathbf{R}_{\text{c.m.}} \approx \frac{\sum_k \left(m_k c^2 + \frac{1}{2} m_k v_k^2 + e_k \sum_i' \frac{e_i}{|\mathbf{r}_k - \mathbf{r}_i|} \right) \mathbf{r}_k}{E_0 + E_1}. \quad (25.25)$$

One can convince oneself of this by a direct verification of the equality

$$\mathbf{V}_{\text{c.m.}} = \frac{d\mathbf{R}_{\text{c.m.}}}{dt} = \frac{\mathbf{P}c^2}{E_{\text{tot}}},$$

which is valid with an accuracy up to quantities of the order of v^2/c^2 .

In addition to the energy and momentum, a system of material points possesses angular momentum

$$\mathbf{L} = \sum_k \mathbf{r}_k \times \frac{\partial \mathbf{L}}{\partial \mathbf{v}_k} = \mathbf{L}_0 + \mathbf{L}_1, \quad (25.26)$$

where \mathbf{L}_0 is the angular momentum of classical mechanics, and \mathbf{L}_1 is the relativistic correction depending on the velocities and coordinates of all points of the system.

We see that in this approximation (taking account of the corrections v^2/c^2) in the relativistic mechanics of the system one can introduce the same basic notions as in classical mechanics. However, in this approximation also the system possesses no potential energy.

Thus, in addition to the case of a system of particles interacting via collisions as discussed in § 15, one can construct the general mechanics of a system of interacting charged particles in the theory of relativity. However, in this case the theory has an approximate character and the highest terms retained in it are of the order of v^2/c^2 . The account of subsequent terms of the expansion in powers of (v/c) is possible only on concrete systems possessing no dipole radiation (the terms $(v/c)^3$), no quadrupole radiation $((v/c)^4)$ and so on.

In what follows we shall need another representation of the energy of interaction in the case where the system consists of two particles. Writing the Lagrangian of the first particle in the form

$$L_1 = -m_1 c^2 \sqrt{1 - \frac{v_1^2}{c^2}} - e_1 \varphi_1 + \frac{e_1}{c} (\mathbf{v}_1 \cdot \mathbf{A}_1),$$

we shall consider φ_1 and A_1 to be the potentials of the field produced by the second particle at the point where the first particle is located at an instant t . Taking account of the retardation and for an arbitrary law of motion the potentials φ_1 and A_1 represent Liénard-Wiechert potentials. The Liénard-Wiechert potentials φ_1 and A_1 are related by formula (25.5) of Part I:

$$A_1 = \frac{\mathbf{v}_2 \varphi_1}{c}.$$

Hence

$$L_1 = -m_1 c^2 \sqrt{1 - \frac{v_1^2}{c^2}} - e \left(1 - \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{c^2} \right) \varphi_1.$$

Thus it follows that for the energy of interaction of two particles one can write the expression

$$U_{\text{interac}} = e \left(1 - \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{c^2} \right) \varphi, \quad (25.27)$$

where φ is the field potential depending only on the instantaneous distance $R(\tau)$ between the charges.

§26. The radiation emitted by a moving charge

Formula (28.4) of Part I for the radiation emitted by a moving charge is applicable only at velocities which are small in comparison with the velocity of light. In order to obtain an analogous expression valid at velocities close to the velocity of light, we introduce into the treatment a set of accompanying reference frames in one of which the particle is at rest at each instant. In every one of these reference frames the formula (28.4) of Part I for the radiation is valid. The radiation described by this formula has the character of spherical waves, so that the total momentum of the emitted electromagnetic waves is equal to zero. The energy emitted per unit time by a charge is, according to (22.16), an invariant:

$$-\frac{dE}{dt} = -\frac{dE'}{dt'} = \dots = \text{invar}. \quad (26.1)$$

The rate of loss of energy due to radiation can be written, according to (28.4) of Part I,

$$\frac{dE'}{dt'} = -\frac{2e^2}{3c^3} (w')^2 = -\frac{2e^2}{3c^2} (w'_\alpha)^2, \quad (26.2)$$

since in the accompanying reference frame $w'_\tau = 0$. The rate of loss of momentum due to radiation is, according to (28.5) of Part I, equal to zero:

$$\frac{dp'}{dt'} = 0. \quad (26.3)$$

To find the radiation in an arbitrary (unprimed) reference frame one need only transform the square of the acceleration $(w'_\alpha)^2$ according to formula (11.17) to the acceleration in the unprimed reference frame.

We then have

$$\frac{dE}{dt} = \frac{dE'}{dt'} = -\frac{2e^2}{3c^3} \frac{\dot{v}^2 - \left[\frac{\mathbf{v} \times \dot{\mathbf{v}}}{c} \right]^2}{\left(1 - \frac{v^2}{c^2} \right)^3} = -\frac{2e^2}{3c^3} \frac{\dot{v}^2 \left(1 - \frac{v^2}{c^2} \right) + \frac{1}{c^2} (\mathbf{v} \cdot \dot{\mathbf{v}})^2}{\left(1 - \frac{v^2}{c^2} \right)^3} \quad (26.4)$$

In this case we obtain for the change in the momentum per unit time using (13.6) and (26.4)

$$\frac{d\mathbf{p}}{dt} = \frac{\mathbf{v}}{c^2} \frac{dE'}{dt'} \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = -\frac{2e^2}{3c^3} \frac{\dot{v}^2 \left(1 - \frac{v^2}{c^2} \right) + \frac{1}{c^2} (\mathbf{v} \cdot \dot{\mathbf{v}})^2}{\left(1 - \frac{v^2}{c^2} \right)^3} \frac{\mathbf{v}}{c^2} \quad (26.5)$$

It is obvious that for $v/c \ll 1$ formulae (26.4) and (26.5) reduce to (26.2) and (26.3).

Formulae (26.4) and (26.5) allow one to find in an arbitrary reference frame, for example the laboratory system, the energy and momentum of the radiation field produced by a charge with an accelerated motion.

As a rule, an accelerated motion of rapidly moving particles is associated with the action of the electromagnetic field on them. To transform formulae (26.4) and (26.5) in this particular case we make use of the expression (12.14) for the acceleration of a particle in an electromagnetic field. Putting in the value of the Lorentz force, we find

$$\mathbf{w} = \dot{\mathbf{v}} = \frac{e}{m} \sqrt{1 - \frac{v^2}{c^2}} \left\{ \mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} - \frac{\mathbf{v}}{c^2} (\mathbf{v} \cdot \mathbf{E}) \right\} \quad (26.6)$$

Making use of this expression for the acceleration, we have

$$\begin{aligned} \dot{v}^2 \left(1 - \frac{v^2}{c^2}\right) + \frac{1}{c^2} (\mathbf{v} \cdot \dot{\mathbf{v}})^2 &= \frac{e^2}{m^2} \left(1 - \frac{v^2}{c^2}\right)^2 \left\{ \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H}\right)^2 + \right. \\ &\quad \left. + \frac{v^2}{c^4} (\mathbf{v} \cdot \mathbf{E})^2 - \frac{2(\mathbf{v} \cdot \mathbf{E})^2}{c^2} \right\} + \frac{e^2}{m^2} \left(1 - \frac{v^2}{c^2}\right)^3 \frac{(\mathbf{v} \cdot \mathbf{E})^2}{c^2} = \\ &= \frac{e^2}{m^2} \left(1 - \frac{v^2}{c^2}\right)^2 \left\{ \left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{H}}{c}\right)^2 - \frac{(\mathbf{v} \cdot \mathbf{E})^2}{c^2} \right\}. \quad (26.7) \end{aligned}$$

The rate of emission of energy by a charge moving in an electromagnetic field is equal to

$$\begin{aligned} \frac{dE}{dt} &= - \frac{2e^4}{3m^2c^3} \frac{\left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H}\right)^2 - \frac{1}{c^2} (\mathbf{v} \cdot \mathbf{E})^2}{1 - \frac{v^2}{c^2}} = \\ &= - \frac{2e^4}{3m^4c^7} \left[\frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} \right]^2 \left\{ \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H}\right)^2 - \frac{1}{c^2} (\mathbf{v} \cdot \mathbf{E})^2 \right\} = \\ &= - \frac{2e^4}{3m^4c^7} E^2 \left\{ \left(\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H}\right)^2 - \frac{1}{c^2} (\mathbf{v} \cdot \mathbf{E})^2 \right\}. \quad (26.8) \end{aligned}$$

Consider several cases of formula (26.8) for the ultrarelativistic case $v \approx c$. Let there be an electric field only (i.e. $\mathbf{H} = 0$). Then

$$\frac{1}{E^2} \frac{dE}{dt} = - \frac{2e^4}{3m^4c^7} \left\{ (\mathbf{E})^2 - \frac{(\mathbf{v} \cdot \mathbf{E})^2}{c^2} \right\}. \quad (26.9)$$

For $\mathbf{v} \perp \mathbf{E}$ the rate of energy loss is

$$\frac{dE}{dt} = - \frac{2e^4}{3m^4c^7} (\mathbf{E})^2 E^2 \quad (26.10)$$

For $\mathbf{v} \parallel \mathbf{E}$ the rate of energy loss is

$$\frac{dE}{dt} = -\frac{2e^4 E^2}{3m^4 c^7} (E)^2 \left(1 - \frac{v^2}{c^2}\right) = -\frac{2e^4}{3m^2 c^3} (E)^2, \quad (26.11)$$

and does not depend on the energy.

For the motion in a magnetic field perpendicular to the direction of the velocity ($\mathbf{v} \perp \mathbf{H}$) and $\mathbf{E} = 0$ we find

$$\frac{dE}{dt} = -\frac{2e^4}{3m^2 c^5} \frac{v^2 (\mathbf{H})^2}{1 - \frac{v^2}{c^2}} = -\frac{2e^4}{3m^4 c^5} p^2 (\mathbf{H})^2 \approx -\frac{2e^4}{3m^4 c^7} E^2 (\mathbf{H})^2, \quad (26.12)$$

where p is the momentum of the particle.

Formulae (26.9)–(26.12) are used in nuclear physics for the determination of energy losses by ultrarelativistic particles moving in electric and magnetic fields. The motion of charged particles in cosmic rays in the Earth's magnetic field and in the magnetic field of a betatron are examples of the motion of ultrarelativistic particles in a magnetic field. Calculations have shown that energy losses due to radiation in a magnetic field determine the upper limit of the energy of particles which can reach the Earth's surface, as well as the upper limit of energies to which electrons can be accelerated in a betatron.

An important application of the formulae obtained is in the calculation of the bremsstrahlung of ultrarelativistic particles in the electric field of a nucleus. An ultrarelativistic electron passing by a nucleus undergoes a very small deflection. Its velocity can be assumed to be constant, and the acceleration can be assumed to be perpendicular to the direction of velocity and equal to $w_{\perp} = eE_{\perp}/m$, where

$$E_{\perp} = Ze\rho/r^3$$

is the component of the nuclear field perpendicular to the velocity (we choose the direction of the latter as the x -axis). For the transverse acceleration use can be made of the non-relativistic expression, since the corresponding velocity component is very small. As in §43 in Part I, ρ is the impact parameter, and r is the distance between the nucleus and the electron. For a motion with constant velocity it can be assumed that

$$r = (\rho^2 + v^2 t^2)^{1/2}$$

Formula (26.10) gives for the rate of energy loss

$$\begin{aligned}\frac{dE}{dt} &= -\frac{2e^2}{3m^2c^3} \frac{1}{1-\frac{v^2}{c^2}} (E_{\perp})^2 = \\ &= -\frac{2Z^2e^4}{3m^2c^3} \frac{1}{1-\frac{v^2}{c^2}} \frac{\rho^2}{(\rho^2 + v^2t^2)^3}.\end{aligned}$$

Integrating with respect to the transit time, we obtain the total energy loss due to bremsstrahlung by an ultrarelativistic particle:

$$\Delta E = -\frac{2Z^2e^4}{3m^2c^3} \frac{\rho^2}{\left(1-\frac{v^2}{c^2}\right)} \int_{-\infty}^{\infty} \frac{dt}{(\rho^2 + v^2t^2)^3}.$$

We have, obviously,

$$\int_{-\infty}^{\infty} \frac{dt}{\left(1 + \frac{v^2t^2}{\rho^2}\right)^3} = 2 \int_0^{\infty} \frac{dt}{\left(1 + \frac{v^2}{\rho^2}t^2\right)^3} = 2\frac{\rho}{v} \int_0^{\pi/2} \cos^4 z \, dz = \frac{3}{8} \frac{\pi\rho}{v},$$

Hence

$$\Delta E = -\frac{\pi}{4} \frac{Z^2e^4}{m^2c^3v} \frac{1}{\left(1-\frac{v^2}{c^2}\right)} \frac{1}{\rho^3}. \quad (26.13)$$

The magnitude of the energy loss increases rapidly with increasing atomic number Z of the matter in which the particle is moving. Formula (26.13) determines the energy loss of one particle passing by a nucleus at a distance ρ . It shows that the loss increases rapidly with decreasing ρ .

In practice the particle may pass by the nucleus at any distance. Multiplying (26.13) by $2\pi\rho \, d\rho n$, where n is the density of the beam, and integrating over all values of ρ , we find the effective radiation of a beam of particles:

$$\begin{aligned}
 E_{\text{eff}} &= -\frac{\pi^2}{2} \frac{Z^2 e^4}{m^2 c^3 v} \frac{n}{\left(1 - \frac{v^2}{c^2}\right)} \int_{\rho_{\min}}^{\infty} \frac{d\rho}{\rho^2} = \\
 &= -\frac{\pi^2}{2} \frac{Z^2 e^4}{m^2 c^3 v} \frac{n}{\left(1 - \frac{v^2}{c^2}\right)} \frac{1}{\rho_{\min}}.
 \end{aligned} \tag{26.14}$$

In formula (26.14) the closest distance of approach of the electron to the nucleus, ρ_{\min} , is introduced, since the integral diverges at the lower limit. The introduction of this unknown quantity means that the classical theory of radiation turns out to be inapplicable for the calculation of bremsstrahlung.

In quantum mechanics it will be shown that the classical treatment of the motion of the electron is inapplicable at small distances. A quantum-mechanical calculation leads to the value

$$\rho_{\min} = \frac{h}{mc} \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}},$$

so that

$$E_{\text{eff}} = -\frac{\pi^2}{2} \frac{Z^2 e^4}{mc^2 v h} \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{26.15}$$

Formula (26.15) allows one to find the losses due to bremsstrahlung when very fast particles pass through matter.

The comparison of (26.15) with the formula for energy loss due to ionization shows that bremsstrahlung is the basic factor determining the deceleration of fast electrons in matter. Losses due to the bremsstrahlung are important for electrons at energies of the order of $200 mc^2$ (100 MeV) in air and $20 mc^2$ (10 MeV) in lead. For heavy particles, for example protons, almost all losses are associated with the ionization up to very large energies.

In conclusion it should be noted that one cannot pass directly from formula (26.14) to the non-relativistic formula (43.37) of Part I, assuming $v \ll c$. Formula (26.14) is found for $v \approx c$, and not for the general case of an arbitrary velocity.

Vector Analysis

Assuming that the reader is familiar with vector analysis, we summarize below the basic formulae used in this book.

Vector algebra

$$\mathbf{a} = a_x \mathbf{i} + a_y \mathbf{j} + a_z \mathbf{k} = a_0 \mathbf{a}_0,$$

where \mathbf{i} , \mathbf{j} , \mathbf{k} are unit vectors directed along the x -axis, y -axis and z -axis respectively, and \mathbf{a}_0 is the unit vector in the direction of \mathbf{a} .

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} = ab \cos(\widehat{\mathbf{a}, \mathbf{b}}) = a_x b_x + a_y b_y + a_z b_z; \quad (1.1)$$

$$\begin{aligned} \mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} = \\ &= (a_y b_z - a_z b_y) \mathbf{i} + (a_z b_x - a_x b_z) \mathbf{j} + (a_x b_y - a_y b_x) \mathbf{k}, \end{aligned} \quad (1.2)$$

$$|\mathbf{a} \times \mathbf{b}| = ab \sin(\widehat{\mathbf{a}, \mathbf{b}}), \quad (1.2')$$

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}), \quad (1.3)$$

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b}), \quad (1.4a)$$

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = \mathbf{b}(\mathbf{c} \cdot \mathbf{a}) - \mathbf{a}(\mathbf{c} \cdot \mathbf{b}) \quad (1.4b)$$

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}). \quad (1.5)$$

$$\begin{aligned} \mathbf{a}(\mathbf{b} \cdot \mathbf{c}) &= \frac{1}{2} \{ \mathbf{a}(\mathbf{b} \cdot \mathbf{c}) - \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) \} + \frac{1}{2} \{ \mathbf{a}(\mathbf{b} \cdot \mathbf{c}) + \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) \} = \\ &= \frac{1}{2} \mathbf{c} \times (\mathbf{a} \times \mathbf{b}) + \frac{1}{2} \{ \mathbf{a}(\mathbf{b} \cdot \mathbf{c}) + \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) \}. \end{aligned} \quad (1.6)$$

The basic formula of spherical trigonometry is easily derived from (1.5). Let \mathbf{r}_1 , \mathbf{r}_2 and \mathbf{r}_3 be the unit radius vectors of the apices of a spherical triangle ABC with angles α , β , γ as in fig. A.1. If in (1.5) we set $\mathbf{a} = \mathbf{r}_1$, $\mathbf{b} = \mathbf{r}_2$, $\mathbf{c} = \mathbf{r}_3$, $\mathbf{d} = \mathbf{r}_1$, we find

$$(\mathbf{r}_1 \times \mathbf{r}_2) \cdot (\mathbf{r}_1 \times \mathbf{r}_3) = \mathbf{r}_2 \cdot \mathbf{r}_3 - (\mathbf{r}_1 \cdot \mathbf{r}_2)(\mathbf{r}_1 \cdot \mathbf{r}_3). \quad (1.5')$$

By definition

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = \cos \gamma, \quad \mathbf{r}_2 \cdot \mathbf{r}_3 = \cos \alpha, \quad \mathbf{r}_3 \cdot \mathbf{r}_1 = \cos \beta,$$

$$|\mathbf{r}_1 \times \mathbf{r}_2| = \sin \gamma, \quad |\mathbf{r}_1 \times \mathbf{r}_3| = \sin \beta,$$

$$(\mathbf{r}_1 \times \mathbf{r}_3) \cdot (\mathbf{r}_1 \times \mathbf{r}_2) = \sin \beta \sin \gamma \cos \delta,$$

where δ is the angle between the planes OAB and OAC.

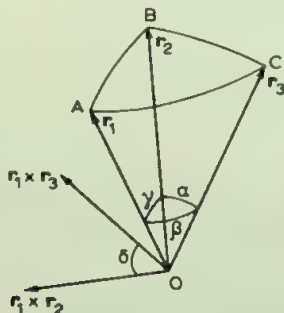


Fig. A.1

The substitution of these expressions into (I.5') gives the basic formula of spherical trigonometry:

$$\cos \alpha = \cos \beta \cos \gamma + \sin \beta \sin \gamma \cos \delta \quad (1.7)$$

Vectors which do not change under the inversion of the coordinate axes $\mathbf{r} \rightarrow (-\mathbf{r})$ are called polar vectors. The velocity vector, force vector etc. are polar vectors. Vectors which change sign under the inversion, i.e.

$$\mathbf{a}(\mathbf{r}) = -\mathbf{a}(-\mathbf{r})$$

are called axial vectors or pseudo-vectors.

The vector expressing the cross product of two polar vectors is an axial vector.

Scalar field

A scalar field $\varphi(\mathbf{r})$ is characterized by defining the scalar φ at every point of space. The spatial rate of change of the scalar $\varphi(\mathbf{r})$ is characterized by the derivative with respect to a given direction \mathbf{l} :

$$\begin{aligned} \frac{\partial \varphi}{\partial l} &= \frac{\partial \varphi}{\partial x} \cos(\widehat{\mathbf{l}, \mathbf{i}}) + \frac{\partial \varphi}{\partial y} \cos(\widehat{\mathbf{l}, \mathbf{j}}) + \frac{\partial \varphi}{\partial z} \cos(\widehat{\mathbf{l}, \mathbf{k}}) \\ &= |\mathbf{l}| |\nabla \varphi| \cos(\widehat{\mathbf{l}, \nabla \varphi}), \end{aligned} \quad (1.8)$$

where the gradient of the scalar φ ,

$$\nabla \varphi = \frac{\partial \varphi}{\partial x} \mathbf{i} + \frac{\partial \varphi}{\partial y} \mathbf{j} + \frac{\partial \varphi}{\partial z} \mathbf{k}, \quad (1.9)$$

represents the vector which is oriented in the direction of the most rapid increase of φ and is equal to the derivative of φ with respect to this direction.

The magnitude of the gradient is equal to

$$|\nabla \varphi| = \sqrt{\left(\frac{\partial \varphi}{\partial x}\right)^2 + \left(\frac{\partial \varphi}{\partial y}\right)^2 + \left(\frac{\partial \varphi}{\partial z}\right)^2} \quad (1.10)$$

The Hamiltonian differential operator "del" ∇ is defined by the relation

$$\nabla \equiv \mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} + \mathbf{k} \frac{\partial}{\partial z}. \quad (\text{I.11})$$

∇ is not represented as a vector. The Hamiltonian operator is a symbol indicating operations which must be done on the functions of coordinates standing in front of it. For example, $\nabla \varphi$ means that one has to take the partial derivatives of the function φ and construct a vector whose components are equal to these partial derivatives.

From the definition (I.11) it follows that

$$\nabla r = \mathbf{i} \frac{\partial r}{\partial x} + \mathbf{j} \frac{\partial r}{\partial y} + \mathbf{k} \frac{\partial r}{\partial z} = \frac{\mathbf{r}}{r}. \quad (\text{I.12})$$

$$\nabla \varphi(r) = \frac{d\varphi}{dr} \nabla r = \frac{d\varphi}{dr} \frac{\mathbf{r}}{r}, \quad (\text{I.13})$$

$$\nabla (\varphi + \psi) = \nabla \varphi + \nabla \psi, \quad (\text{I.14})$$

$$\nabla \varphi \psi = \psi \nabla \varphi + \varphi \nabla \psi, \quad (\text{I.15})$$

$$\nabla f(\varphi) = \frac{df}{d\varphi} \nabla \varphi. \quad (\text{I.16})$$

In calculating the gradient of functions which depend on the distance r between two given points,

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2},$$

one has to distinguish between the gradients with respect to coordinates (x, y, z) and (x_0, y_0, z_0) .

We have

$$\nabla r = \frac{\mathbf{i}(x - x_0) + \mathbf{j}(y - y_0) + \mathbf{k}(z - z_0)}{r}, \quad (\text{I.12}')$$

$$\nabla_0 r = - \frac{\mathbf{i}(x - x_0) + \mathbf{j}(y - y_0) + \mathbf{k}(z - z_0)}{r}.$$

Consequently,

$$\nabla_0 \varphi(r) = - \nabla \varphi(r) . \quad (I.17)$$

The differentiation of a vector which depends on a scalar argument:

$$\frac{da(x)}{dx} = \frac{d}{dx} a_0(x) a(x) = a_0 \frac{da}{dx} + a \frac{da_0}{dx} = a_0 \frac{da}{dx} + b ,$$

where $a_0(x) = a(x)/|a(x)|$, $b = a\omega b_0$, $b_0 \perp a_0$, and $\omega = |da_0/dx|$ is the rate of change of the angle φ , which determines the orientation of the vector a .

The total derivative of $A(x, y, z, t)$ with respect to time is equal to

$$\begin{aligned} \frac{dA}{dt} &= \frac{\partial A}{\partial t} + \frac{\partial A}{\partial x} \frac{dx}{dt} + \frac{\partial A}{\partial y} \frac{dy}{dt} + \frac{\partial A}{\partial z} \frac{dz}{dt} = \\ &= \frac{\partial A}{\partial t} + v_x \frac{\partial A}{\partial x} + v_y \frac{\partial A}{\partial y} + v_z \frac{\partial A}{\partial z} = \frac{\partial A}{\partial t} + (\mathbf{v} \cdot \nabla) A . \end{aligned} \quad (I.18)$$

The line integral of a vector along a path is defined by the relation

$$\int_L \mathbf{a} \cdot d\mathbf{l} = \int (a_x dl_x + a_y dl_y + a_z dl_z) .$$

The integral over a closed path is called the circulation:

$$\oint \mathbf{a} \cdot d\mathbf{l} .$$

If the vector \mathbf{a} can be written in the form

$$\mathbf{a} = \nabla \varphi ,$$

then it is called a potential, or irrotational, vector, and φ is called the potential.

The circulation of a potential vector is equal to zero:

$$\oint \nabla \varphi \cdot d\mathbf{l} = 0 . \quad (I.19)$$

The line integral of a potential vector along the path L between two points \mathbf{r}_1 and \mathbf{r}_2 is

$$\int_L \nabla \varphi \cdot d\mathbf{l} = \varphi(\mathbf{r}_1) - \varphi(\mathbf{r}_2), \quad (1.20)$$

where $\varphi(\mathbf{r}_1)$ and $\varphi(\mathbf{r}_2)$ are the potentials at the end points of the line L .

The scalar field is represented geometrically by plotting equipotential surfaces $\varphi = \text{const}$. The vector $\nabla \varphi$ is normal to the surface $\varphi = \text{const}$. The more rapid the variation of the function φ , the larger $\nabla \varphi$ and the smaller the separation between the equipotential surfaces.

The integral

$$\int \varphi(\mathbf{r}) dS = \int \varphi \mathbf{n} dS = \lim_{\substack{N \rightarrow \infty \\ \Delta S_i \rightarrow 0}} \sum_{i=1}^N \varphi_i \Delta S_i$$

is called the surface integral of the function $\varphi(\mathbf{r})$. The integral over a closed surface is denoted in this book by $\oint \varphi dS$.

Let us consider the integral over the surface of an infinitesimal parallelepiped of volume $V \rightarrow 0$. We assume that the faces of the parallelepiped are infinitesimal areas $(dx dy)$, $(dx dz)$, $(dy dz)$ in the coordinate planes (xy) , (xz) and (yz) at the origin.

We have, obviously,

$$\begin{aligned} \oint_{V \rightarrow 0} \varphi dS &= i \{ \varphi(0 + dx) dy dz - \varphi(0) dy dz \} + \\ &+ j \{ \varphi(0 + dy) dx dz - \varphi(0) dx dz \} + \\ &+ k \{ \varphi(0 + dz) dx dy - \varphi(0) dx dy \} = \\ &= \left(i \frac{\partial \varphi}{\partial x} + j \frac{\partial \varphi}{\partial y} + k \frac{\partial \varphi}{\partial z} \right) dx dy dz \\ &= \left(i \frac{\partial \varphi}{\partial x} + j \frac{\partial \varphi}{\partial y} + k \frac{\partial \varphi}{\partial z} \right) V, \end{aligned}$$

whence we find the equality

$$\nabla \varphi = i \frac{\partial \varphi}{\partial x} + j \frac{\partial \varphi}{\partial y} + k \frac{\partial \varphi}{\partial z} = \lim_{V \rightarrow 0} \frac{\oint \varphi \mathbf{n} dS}{V} \quad (1.21)$$

Formula (I.21) allows one to give another, integral, definition of the Hamiltonian operator:

$$\nabla = \lim_{V \rightarrow 0} \frac{\oint \mathbf{n} dS}{V}, \quad (I.22)$$

which is equivalent with (I.11).

We emphasize that, since the integration surface reduces to a point as V tends to zero, the integral operator (I.22) does not depend on the form of this surface.

The following integral relation results from (I.21):

$$\oint \varphi dS = \int \nabla \varphi dV. \quad (I.23)$$

This relation connects the surface integral of the scalar φ with the volume integral of the vector $\nabla \varphi$. The volume of integration on the right-hand side of (I.23) is bounded by the surface φ over which the surface integration on the left-hand side of (I.23) is carried out.

In order to prove formula (I.23), we divide the finite volume into infinitesimal volumes, formula (I.21) being valid for each of these. We carry out the summation over all these volumes. The integration over all internal surfaces which represent interfaces of the volumes is carried out twice. In this case the directions of the external normals will be opposite, and the integrals over the internal surfaces will cancel out. There will remain only the integrals over all external surfaces, which form in the sum the integral over the surface bounding the volume V .

Vector field

The region of space in which the value of a vector $\mathbf{a}(\mathbf{r})$ is defined at each point is called a vector field. The vector field is represented graphically by means of field lines. The vector $\mathbf{a}(\mathbf{r}_0)$ is in the direction of a tangent to the field line at the point \mathbf{r}_0 . The field lines are drawn in such a way that the density of lines is proportional to the absolute value $|\mathbf{a}|$.

For the vector field one can define the flux of the vector through an area characterized by the vector $d\mathbf{S}$ in the form

$$dJ = \mathbf{a} \cdot d\mathbf{S}. \quad (I.24)$$

By the surface integral of a vector \mathbf{a} we mean the quantity

$$\begin{aligned} j &= \int_S \mathbf{a} \cdot d\mathbf{S} = \int_S \mathbf{a} \cdot \mathbf{n} dS = \int_S a_n dS \\ &= \int a_x dy dz + \int a_y dx dz + \int a_z dx dy, \end{aligned} \quad (1.25)$$

where $dy dz = dS \cos(\widehat{\mathbf{n}}, \widehat{\mathbf{i}})$, and so on.

The surface integral represents the flux of the vector \mathbf{a} through the surface S .

If the surface S is closed, then the surface integral is denoted by $\oint \mathbf{a} \cdot d\mathbf{S}$. For the surface integral over a closed surface the Gauss-Ostrogradsky theorem

$$\oint \mathbf{a} \cdot d\mathbf{S} = \int \left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} \right) dV \quad (1.26)$$

holds.

The integration in (1.26) is carried out over the volume bounded by the surface of integration of the surface integral.

The proof is obtained by dividing the volume into infinitesimal volumes. It can formally be performed by means of the definition (1.22) of the Hamiltonian operator.

Namely,

$$\nabla \cdot \mathbf{a} = \lim_{V \rightarrow 0} \frac{\oint \mathbf{a} \cdot \mathbf{n} dS}{V}, \quad (1.27)$$

whence, as in deriving (1.23), summing the elementary volumes we obtain

$$\int \nabla \cdot \mathbf{a} dV = \oint \mathbf{a} \cdot d\mathbf{S}.$$

Making use of the definition (1.11), we arrive at theorem (1.26).

The scalar quantity which may be expressed by two equivalent representations

$$\nabla \cdot \mathbf{a} = \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}, \quad (1.28)$$

$$\nabla \cdot \mathbf{a} = \lim_{V \rightarrow 0} \frac{\oint \mathbf{a} \cdot \mathbf{n} dS}{V}, \quad (1.29)$$

which are based on different forms of the Hamiltonian operator is called the divergence of the vector \mathbf{a} . The Gauss-Ostrogradsky theorem can be rewritten in the form

$$\oint \mathbf{a} \cdot d\mathbf{S} = \int \nabla \cdot \mathbf{a} dV. \quad (1.30)$$

The divergence plays an important role in the theory of the vector field (see §2 of Part I of the book).

From expression (1.29) it follows that $\nabla \cdot \mathbf{a}$ represents the flux per unit volume of the vector $\mathbf{a}(\mathbf{r})$ through an infinitesimal surface surrounding the given point \mathbf{r} of the field.

A vector field is called solenoidal, if at each point of the field

$$\nabla \cdot \mathbf{a} = 0.$$

This means that the flux of the vector through a transverse section of a tube formed by a group of field lines has a constant value along the tube. At those points of the field at which $\nabla \cdot \mathbf{a} \neq 0$ there are sources ($\nabla \cdot \mathbf{a} > 0$) or sinks ($\nabla \cdot \mathbf{a} < 0$) of the field. The numerical value of $\nabla \cdot \mathbf{a}$ is called the intensity or abundance of sources of the field.

The following obvious formulae hold:

$$\begin{aligned} \nabla \cdot (\mathbf{a}_1 + \mathbf{a}_2) &= \nabla \cdot \mathbf{a}_1 + \nabla \cdot \mathbf{a}_2, \\ \nabla \cdot (c\mathbf{a}) &= c(\nabla \cdot \mathbf{a}) \quad (c = \text{const}), \\ \nabla \cdot \mathbf{a}(\mathbf{r}) &= -\nabla_0 \cdot \mathbf{a}(\mathbf{r}), \end{aligned} \quad (1.31)$$

where $r = \sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2}$; and the index 0 denotes differentiation with respect to the coordinates x_0, y_0, z_0 .

Let $\mathbf{a}(u)$ be a vector depending only on the scalar quantity u . Then

$$\nabla \cdot \mathbf{a}(u) = \frac{d\mathbf{a}}{du} \cdot \nabla u = \dot{\mathbf{a}} \cdot \nabla u, \quad (1.32)$$

where the dot over \mathbf{a} denotes differentiation with respect to the argument u .

Besides the operation $\nabla \cdot \mathbf{a}$ corresponding to the scalar product of the vectors ∇ and \mathbf{a} , one can consider the operation forming the vector

$$\nabla \times \mathbf{a} \equiv \text{curl } \mathbf{a}. \quad (1.33)$$

The vector representing the cross product of the Hamiltonian operator ∇ and the vector \mathbf{a} is called the curl of the vector \mathbf{a} .

A calculation according to formula (1.2), taking into account (1.11), gives

$$\begin{aligned}\nabla \times \mathbf{a} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ a_x & a_y & a_z \end{vmatrix} = \\ &= \mathbf{i} \left(\frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z} \right) + \mathbf{j} \left(\frac{\partial a_x}{\partial z} - \frac{\partial a_z}{\partial x} \right) + \mathbf{k} \left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y} \right). \quad (1.34)\end{aligned}$$

Making use now of the integral expression (1.22) for the Hamiltonian operator, we have another representation of the vector:

$$\nabla \times \mathbf{a} = \lim_{V \rightarrow 0} \frac{\oint \mathbf{n} \times \mathbf{a} dS}{V}. \quad (1.35)$$

From the definition (1.35), in an analogous way to (1.23) and (1.30), it follows that

$$\int \nabla \times \mathbf{a} dV = \oint \mathbf{dS} \times \mathbf{a}. \quad (1.36)$$

Let us now consider the projection of the vector $\nabla \times \mathbf{a}$ on to an arbitrary direction characterized by unit vector \mathbf{N} . From the definition (1.35) it follows that

$$\mathbf{N} \cdot (\nabla \times \mathbf{a}) = \lim_{V \rightarrow 0} \frac{\mathbf{N} \cdot \oint (\mathbf{n} \times \mathbf{a}) dS}{V} = \lim_{V \rightarrow 0} \frac{\oint \mathbf{a} \cdot (\mathbf{N} \times \mathbf{n}) dS}{V}.$$

Since the result obtained after passing to the limit does not depend on the form of the surface, the latter can be chosen arbitrarily. If we direct the z -axis along \mathbf{N} and choose a cylinder with base S and height h as the surface, then

$$\oint \mathbf{a} \cdot (\mathbf{N} \times \mathbf{n}) dS = \oint \mathbf{a} \cdot d\mathbf{l} h.$$

The integration in the expression on the right is carried out with respect to the lateral surface of the cylinder, since at its bases $\mathbf{n} \parallel \mathbf{N}$ and their cross product reduces to zero. A length element at the lateral surface perpendicular to the vectors \mathbf{n} and \mathbf{N} is denoted by $d\mathbf{l}$.

Hence it follows that

$$\bar{\mathbf{N}} \cdot (\nabla \times \mathbf{a}) = \lim_{V \rightarrow 0} \frac{h}{Sh} \oint \mathbf{a} \cdot d\mathbf{l} = \lim_{S \rightarrow 0} \frac{1}{S} \oint \mathbf{a} \cdot d\mathbf{l}.$$

Dividing an arbitrary volume into small volumes as previously and carrying out the summation over them, we find that the surface integrals over the internal surfaces and the line integrals over the edges of adjacent cells cancel out, so that

$$\oint \mathbf{a} \cdot d\mathbf{l} = \int (\nabla \times \mathbf{a}) \cdot d\mathbf{S}. \quad (1.37)$$

Formula (1.37), connecting the line integral of the vector \mathbf{a} over an arbitrary closed contour with the surface integral of $\nabla \times \mathbf{a}$, is called Stokes' theorem.

From the derivation of the theorem it is clear that the integration is carried out over an arbitrary surface bounded by the contour of integration.

The geometrical meaning of the idea of curl becomes clear from Stokes' theorem. In order that the integral of the vector \mathbf{a} over a closed path be different from zero it is necessary that certain lines (even field lines) have the character of closed curves. Such lines, for example, are the lines of the vector of the velocity of rotation of a solid body, or the flow lines of a liquid performing a circulatory motion. Hence the term curl.

From Stokes' theorem it follows that:

1) If $\mathbf{a} = \nabla \varphi$, then $\oint \mathbf{a} \cdot d\mathbf{l} = 0$. Consequently,

$$\nabla \times \mathbf{a} = \nabla \times \nabla \varphi = 0, \quad (1.38)$$

i.e. if \mathbf{a} is a potential vector, then the field of the vector \mathbf{a} is irrotational. Conversely, any vector whose field is irrotational is a potential vector.

2) If the vector \mathbf{a} is solenoidal, so that

$$\nabla \cdot \mathbf{a} = 0,$$

then it can be written in the form of the curl of a certain vector \mathbf{c} :

$$\mathbf{a} = \nabla \times \mathbf{c}.$$

Conversely, if a vector field \mathbf{a} can be written as the curl of a field \mathbf{c} , then the field \mathbf{a} has solenoidal character.

$$\nabla \cdot (\nabla \times \mathbf{c}) = 0.$$

Sources and sinks are absent in a solenoidal field. The proof of these statements is obtained by a direct calculation. For example,

$$\nabla \cdot (\nabla \times \mathbf{c}) = \mathbf{c} \cdot (\nabla \times \nabla) = 0.$$

The formation of the scalar $\nabla \cdot \mathbf{a}$ and the vector $\nabla \times \mathbf{a}$ are the basic operations of differentiation of a vector.

The divergence and curl of a vector \mathbf{a} determine the vector field (see §2 of Part I).

In calculating the divergence and curl of a vector $\mathbf{a}(u)$ which depends on the scalar argument u it turns out that

$$\nabla \cdot \mathbf{a}(u) = \nabla u \cdot \frac{d\mathbf{a}}{du} = \nabla u \cdot \dot{\mathbf{a}}, \quad (1.39)$$

$$\nabla \times \mathbf{a}(u) = \nabla u \times \frac{d\mathbf{a}}{du} = (\nabla u) \times \dot{\mathbf{a}}, \quad (1.40)$$

where the dot over \mathbf{a} denotes differentiation with respect to the scalar u .

Calculation of the derivatives of a product and repeated derivatives. These operations are carried out most simply by means of the Hamiltonian operator. In this case the following two rules are to be adhered to:

1. The Hamiltonian operator must act successively on each scalar and vector following it.

2. The Hamiltonian operator can be treated as an ordinary vector, but it cannot be transposed with the quantity on which it acts. For clarity, in carrying out intermediate transformations we shall indicate the quantity acted upon by the Hamiltonian operator by a subscript, for example, ∇_φ or $\nabla_{\mathbf{a}}$.

The most important examples are:

$$1. \quad \nabla(\varphi\psi) = \varphi \nabla_\psi \psi + \psi \nabla_\varphi \varphi = \varphi \nabla \psi + \psi \nabla \varphi. \quad (1.41)$$

$$2. \quad \nabla \cdot (\varphi \mathbf{a}) = \mathbf{a} \cdot \nabla_\varphi \varphi + \varphi \nabla_{\mathbf{a}} \cdot \mathbf{a} = \mathbf{a} \cdot \nabla \varphi + \varphi \nabla \cdot \mathbf{a}. \quad (1.42)$$

In particular

$$\nabla \cdot \left(\frac{\mathbf{r}}{r} \right) \equiv \nabla \cdot \mathbf{n} = \left(\frac{1}{r} \right) \nabla \cdot \mathbf{r} + \mathbf{r} \cdot \nabla \left(\frac{1}{r} \right) = \frac{3}{r} - \frac{1}{r} = \frac{2}{r},$$

$$\nabla \times (\varphi \mathbf{a}) = \varphi (\nabla_{\mathbf{a}} \times \mathbf{a}) + (\nabla_{\varphi} \varphi) \times \mathbf{a} = \varphi (\nabla \times \mathbf{a}) + (\nabla \varphi) \times \mathbf{a}. \quad (1.43)$$

$$\begin{aligned} 3. \quad \nabla \cdot (\mathbf{a} \times \mathbf{b}) &= \nabla_{\mathbf{a}} \cdot (\mathbf{a} \times \mathbf{b}) + \nabla_{\mathbf{b}} \cdot (\mathbf{a} \times \mathbf{b}) = \\ &= \mathbf{b} \cdot (\nabla_{\mathbf{a}} \times \mathbf{a}) - \nabla_{\mathbf{b}} \cdot (\mathbf{b} \times \mathbf{a}) = \\ &= \mathbf{b} \cdot (\nabla_{\mathbf{a}} \times \mathbf{a}) - \mathbf{a} \cdot (\nabla_{\mathbf{b}} \times \mathbf{b}) = \\ &= \mathbf{b} \cdot (\nabla \times \mathbf{a}) - \mathbf{a} \cdot (\nabla \times \mathbf{b}). \end{aligned} \quad (1.44)$$

We have carried out a cyclic permutation of the vectors. In the second term the order of vector multiplication has been changed. Otherwise rule 2 would be violated in the cyclic permutation: the vector \mathbf{b} would be moved behind the symbol $\nabla_{\mathbf{b}}$.

$$\begin{aligned} 4. \quad \nabla \times (\mathbf{a} \times \mathbf{b}) &= \nabla_{\mathbf{a}} \times (\mathbf{a} \times \mathbf{b}) + \nabla_{\mathbf{b}} \times (\mathbf{a} \times \mathbf{b}) = \\ &= (\nabla_{\mathbf{a}} \cdot \mathbf{b})\mathbf{a} - (\nabla_{\mathbf{a}} \cdot \mathbf{a})\mathbf{b} + (\nabla_{\mathbf{b}} \cdot \mathbf{b})\mathbf{a} - (\nabla_{\mathbf{b}} \cdot \mathbf{a})\mathbf{b} = \\ &= (\mathbf{b} \cdot \nabla_{\mathbf{a}})\mathbf{a} - \mathbf{b}(\nabla_{\mathbf{a}} \cdot \mathbf{a}) + \mathbf{a}(\nabla_{\mathbf{b}} \cdot \mathbf{b}) - (\mathbf{a} \cdot \nabla_{\mathbf{b}})\mathbf{b} = \\ &= (\mathbf{b} \cdot \nabla)\mathbf{a} - (\mathbf{a} \cdot \nabla)\mathbf{b} + \mathbf{a}(\nabla \cdot \mathbf{b}) - \mathbf{b}(\nabla \cdot \mathbf{a}). \end{aligned} \quad (1.45)$$

Here $\mathbf{a} \cdot \nabla$ is the scalar differential operator

$$\mathbf{a} \cdot \nabla = a_x \frac{\partial}{\partial x} + a_y \frac{\partial}{\partial y} + a_z \frac{\partial}{\partial z}. \quad (1.46)$$

$$\begin{aligned} 5. \quad \nabla (\mathbf{a} \cdot \mathbf{b}) &= \nabla_{\mathbf{a}} (\mathbf{a} \cdot \mathbf{b}) + \nabla_{\mathbf{b}} (\mathbf{a} \cdot \mathbf{b}) = \\ &= (\mathbf{b} \cdot \nabla_{\mathbf{a}})\mathbf{a} + \mathbf{b} \times (\nabla_{\mathbf{a}} \times \mathbf{a}) + (\mathbf{a} \cdot \nabla_{\mathbf{b}})\mathbf{b} + \mathbf{a} \times (\nabla_{\mathbf{b}} \times \mathbf{b}) = \\ &= (\mathbf{b} \cdot \nabla)\mathbf{a} + (\mathbf{a} \cdot \nabla)\mathbf{b} + \mathbf{b} \times (\nabla \times \mathbf{a}) + \mathbf{a} \times (\nabla \times \mathbf{b}). \end{aligned} \quad (1.47)$$

$$6. \quad \nabla \cdot \frac{1}{2} a^2 = (\mathbf{a} \cdot \nabla)\mathbf{a} + \mathbf{a} \times (\nabla \times \mathbf{a}). \quad (1.48)$$

$$7. \quad \nabla \cdot \nabla \varphi = \nabla^2 \varphi = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \varphi. \quad (1.49)$$

$$\begin{aligned}
 8. \quad \nabla \times (\nabla \times \mathbf{a}) &= \nabla(\nabla \cdot \mathbf{a}) - (\nabla \cdot \nabla) \mathbf{a} = \\
 &= \nabla(\nabla \cdot \mathbf{a}) - \nabla^2 \mathbf{a}.
 \end{aligned} \tag{I.50}$$

$$9. \quad \nabla \cdot (\mathbf{a}\mathbf{b}) = \mathbf{b}(\nabla \cdot \mathbf{a}) + (\mathbf{a} \cdot \nabla) \mathbf{b}. \tag{I.51}$$

The product $(\mathbf{a}\mathbf{b})$ of the two vectors \mathbf{a} and \mathbf{b} is known as dyadic. The properties of dyadics are treated in books on vector analysis.

From the integral representation of the Hamiltonian operator the formula

$$\nabla \cdot (\mathbf{a}\mathbf{b}) = \lim_{V \rightarrow 0} \frac{1}{V} \oint (\mathbf{n} \cdot \mathbf{a}) \mathbf{b} dS$$

follows, from which one obtains directly

$$\int \nabla \cdot (\mathbf{a}\mathbf{b}) dV = \oint (\mathbf{n} \cdot \mathbf{a}) \mathbf{b} dS, \tag{I.52}$$

or by virtue of (I.51),

$$\oint (\mathbf{n} \cdot \mathbf{a}) \mathbf{b} dS = \int \mathbf{b}(\nabla \cdot \mathbf{a}) dV + \int (\mathbf{a} \cdot \nabla) \mathbf{b} dV \tag{I.53}$$

We obtain also another integral equality:

$$\int \mathbf{b} \times (\nabla \times \mathbf{a}) dV + \int (\mathbf{a} \times \nabla) \times \mathbf{b} dV = - \oint (\mathbf{n} \times \mathbf{a}) \times \mathbf{b} dS. \tag{I.54}$$

From the integral representation of the Hamiltonian operator the formula

$$(\nabla \times \mathbf{a}) \times \mathbf{b} = \lim_{V \rightarrow 0} \frac{1}{V} \oint (\mathbf{n} \times \mathbf{a}) \times \mathbf{b} dS$$

follows, whence

$$\int (\nabla \times \mathbf{a}) \times \mathbf{b} dV = \oint (\mathbf{n} \times \mathbf{a}) \times \mathbf{b} dS. \tag{I.55}$$

On the other hand, we have

$$(\nabla \times \mathbf{a}) \times \mathbf{b} = (\nabla_a \times \mathbf{a}) \times \mathbf{b} + (\nabla_b \times \mathbf{a}) \times \mathbf{b} = -\mathbf{b} \times (\nabla \times \mathbf{a}) - (\mathbf{a} \times \nabla) \times \mathbf{b}.$$

Substituting this into (I.55), we arrive at (I.54).

Representation of vector operations in curvilinear coordinates

In addition to Cartesian coordinates it is often convenient to make use of curvilinear coordinates q_1, q_2, q_3 . To each point r there corresponds a set of quantities q_1, q_2 and q_3 , i.e.

$$\mathbf{r} = \mathbf{r}(q_1, q_2, q_3). \quad (1.56)$$

Since vector operations are not connected with any particular coordinate system, the relations of vector analysis remain valid in all coordinate representations. However, the actual expression of vector operations in curvilinear coordinates is, naturally, not the same as that in Cartesian coordinates.

In this book only orthogonal coordinates are used. Let us call the three surfaces

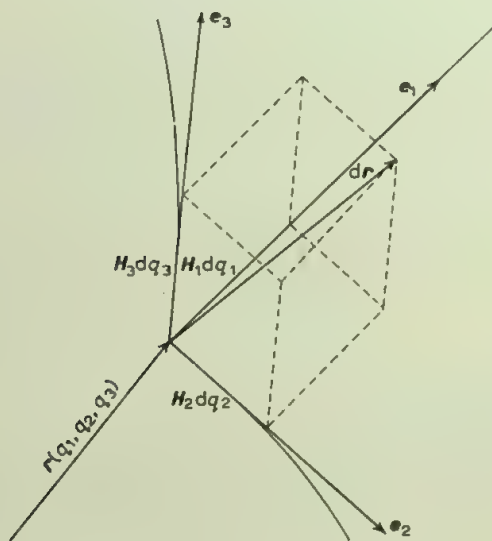


Fig. A.2

$$q_i = q_i(x, y, z) = \text{const} \quad (i = 1, 2, 3), \quad (1.57)$$

the coordinate surfaces, and the lines of their intersection the coordinate lines. It is obvious that the set of three coordinate surfaces is a generalization of the coordinate trihedron in the Cartesian coordinate system. The directions of the coordinate lines are characterized by unit vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ (fig. A.2). Reference frames in which the vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are mutually perpendicular are called orthogonal coordinate systems.

The derivative with respect to coordinates q_i is equal to

$$\frac{\partial \mathbf{r}}{\partial q_i} = \frac{\partial \mathbf{r}(q_1, q_2, q_3)}{\partial q_i} = H_i \mathbf{e}_i \quad (i = 1, 2, 3). \quad (1.58)$$

The vector $\partial \mathbf{r} / \partial q_i$ is directed along a tangent to the coordinate line q_i , and its value H_i is a function of the coordinates q_1, q_2, q_3 . Obviously,

$$H_i^2 = \left(\frac{\partial \mathbf{r}}{\partial q_i} \right)^2 = \left(\frac{\partial x}{\partial q_i} \right)^2 + \left(\frac{\partial y}{\partial q_i} \right)^2 + \left(\frac{\partial z}{\partial q_i} \right)^2 \quad (1.59)$$

The three quantities H_i ($i = 1, 2, 3$) are called the Lamé coefficients. By means of them one can write

$$d\mathbf{r} = H_1 dq_1 \mathbf{e}_1 + H_2 dq_2 \mathbf{e}_2 + H_3 dq_3 \mathbf{e}_3, \quad (1.60)$$

or, in components

$$(d\mathbf{r})_1 = H_1 dq_1; \quad (d\mathbf{r})_2 = H_2 dq_2; \quad (d\mathbf{r})_3 = H_3 dq_3. \quad (1.61)$$

Squaring (1.60), we find the square of the length in orthogonal coordinates

$$(d\mathbf{r})^2 = ds^2 = H_1^2 dq_1^2 + H_2^2 dq_2^2 + H_3^2 dq_3^2 \quad (1.62)$$

An element of area is easily found by considering an infinitesimal parallelepiped formed by the coordinate surfaces. The areas of its faces are

$$\begin{aligned} dS_1 &= H_2 H_3 dq_2 dq_3, & dS_2 &= H_1 H_3 dq_1 dq_3, \\ dS_3 &= H_1 H_2 dq_1 dq_2. \end{aligned} \quad (1.63)$$

The volume of the parallelepiped is equal to

$$dV = H_1 H_2 H_3 dq_1 dq_2 dq_3. \quad (1.64)$$

The most important orthogonal coordinates are cylindrical and spherical polar coordinates.

In spherical polar coordinates

$$q_1 = r, \quad q_2 = \theta, \quad q_3 = \psi. \quad (1.65)$$

Obviously, we have

$$(dr)_r = dr, \quad (dr)_\theta = r d\theta, \quad (dr)_\psi = r \sin \theta d\psi \quad (1.66)$$

Comparison with (1.61) gives

$$\begin{aligned} H_r &= 1, \quad H_\theta = r, \quad H_\psi = r \sin \theta; \\ dV &= r^2 \sin \theta dr d\theta d\psi. \end{aligned} \quad (1.67)$$

In the cylindrical polar coordinate system

$$\begin{aligned} q_1 &= \rho, \quad q_2 = \psi, \quad q_3 = z, \\ (dr)_\rho &= d\rho, \quad (dr)_\psi = \rho d\psi, \quad (dr)_z = dz, \end{aligned} \quad (1.68)$$

and from (1.61)

$$\begin{aligned} H_\rho &= 1, \quad H_\psi = \rho, \quad H_z = 1, \\ dV &= \rho d\rho d\psi dz. \end{aligned} \quad (1.69)$$

Expressions for vector operations are obtained from their definitions, taking into account the relations which have been written out:

$$1. \quad \nabla \varphi = \frac{\mathbf{e}_1}{H_1} \frac{\partial \varphi}{\partial q_1} + \frac{\mathbf{e}_2}{H_2} \frac{\partial \varphi}{\partial q_2} + \frac{\mathbf{e}_3}{H_3} \frac{\partial \varphi}{\partial q_3} \quad (1.70)$$

In particular, in spherical and cylindrical polar coordinate systems we have respectively:

$$\nabla \varphi = \mathbf{e}_r \frac{\partial \varphi}{\partial r} + \frac{\mathbf{e}_\theta}{r} \frac{\partial \varphi}{\partial \theta} + \frac{\mathbf{e}_\psi}{r \sin \theta} \frac{\partial \varphi}{\partial \psi}, \quad (1.71)$$

$$\nabla \varphi = \mathbf{e}_\rho \frac{\partial \varphi}{\partial \rho} + \frac{\mathbf{e}_\psi}{\rho} \frac{\partial \varphi}{\partial \psi} + \mathbf{e}_z \frac{\partial \varphi}{\partial z}. \quad (1.72)$$

$$2. \nabla \cdot \mathbf{a} = \frac{1}{H_1 H_2 H_3} \left\{ \frac{\partial}{\partial q_1} (a_1 H_2 H_3) + \frac{\partial}{\partial q_2} (a_2 H_3 H_1) + \frac{\partial}{\partial q_3} (a_3 H_1 H_2) \right\}. \quad (1.73)$$

In spherical and cylindrical polar coordinates respectively:

$$\nabla \cdot \mathbf{a} = \frac{1}{r^2} \frac{\partial(r^2 a_r)}{\partial r} + \frac{1}{r \sin \theta} \frac{\partial(\sin \theta a_\theta)}{\partial \theta} + \frac{1}{r \sin \theta} \frac{\partial a_\psi}{\partial \psi}, \quad (1.74)$$

$$\nabla \cdot \mathbf{a} = \frac{1}{\rho} \frac{\partial(\rho a_\rho)}{\partial \rho} + \frac{1}{\rho} \frac{\partial a_\psi}{\partial \psi} + \frac{\partial a_z}{\partial z}. \quad (1.75)$$

$$3. (\nabla \times \mathbf{a})_1 = \frac{1}{H_2 H_3} \left\{ \frac{\partial(a_3 H_3)}{\partial q_2} - \frac{\partial(a_2 H_2)}{\partial q_3} \right\}. \quad (1.76)$$

Other projections are obtained by the cyclic permutation of the coordinates 1, 2, 3.

In spherical and cylindrical coordinates respectively:

$$(\nabla \times \mathbf{a})_r = \frac{1}{r \sin \theta} \frac{\partial(a_\psi \sin \theta)}{\partial \theta} - \frac{1}{r \sin \theta} \frac{\partial a_\theta}{\partial \psi}, \quad (1.77)$$

$$(\nabla \times \mathbf{a})_\theta = \frac{1}{r \sin \theta} \frac{\partial a_r}{\partial \psi} - \frac{1}{r} \frac{\partial(r a_\psi)}{\partial r}, \quad (1.78)$$

$$(\nabla \times \mathbf{a})_\psi = \frac{1}{r} \frac{\partial a_\theta r}{\partial r} - \frac{1}{r} \frac{\partial a_r}{\partial \theta}, \quad (1.79)$$

$$(\nabla \times \mathbf{a})_\rho = \frac{1}{\rho} \frac{\partial a_z}{\partial \psi} - \frac{\partial a_\psi}{\partial z}, \quad (1.80)$$

$$(\nabla \times \mathbf{a})_\psi = \frac{\partial a_\rho}{\partial z} - \frac{\partial a_z}{\partial \rho}, \quad (1.81)$$

$$(\nabla \times \mathbf{a})_z = \frac{1}{\rho} \frac{\partial(a_\psi \rho)}{\partial \rho} - \frac{1}{\rho} \frac{\partial a_\rho}{\partial \psi}. \quad (1.82)$$

4. The Laplacian operator is

$$\nabla^2 = \frac{1}{H_1 H_2 H_3} \left\{ \frac{\partial}{\partial q_1} \left(\frac{H_2 H_3}{H_1} \frac{\partial}{\partial q_1} \right) + \frac{\partial}{\partial q_2} \left(\frac{H_3 H_1}{H_2} \frac{\partial}{\partial q_2} \right) + \frac{\partial}{\partial q_3} \left(\frac{H_1 H_2}{H_3} \frac{\partial}{\partial q_3} \right) \right\}. \quad (1.83)$$

In spherical and cylindrical coordinates respectively:

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \psi^2}, \quad (1.84)$$

$$\nabla^2 = \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2}{\partial \psi^2} + \frac{\partial^2}{\partial z^2}. \quad (1.85)$$

Bibliography

- N.E.Kochin, *Vektornoe ischislenie i nachalo tenzornovo ischisleniya (Vector calculus and the principles of tensor calculus)* (Acad. Sci. USSR, 1963).
 Y.I.Frenkel, *Kurs teoreticheskoi mekhaniki (A course of theoretical mechanics)* (Gos-
 tekhnizdat, 1940).
 M.Schwartz, S.Green and W.A.Rutledge, *Vector analysis* (Harper, New York, 1962).

The Fourier Integral

Any function which is periodic in the region $l = 2\pi/\omega$, i.e. a function satisfying the condition

$$f(t + l) = f\left(t + \frac{2\pi}{\omega}\right) = f(t),$$

can be expanded in a Fourier series:

$$f(t) = \sum_{n=0}^{\infty} (a_n \cos n\omega t + b_n \sin n\omega t) \equiv \sum_{n=-\infty}^{\infty} f_n e^{in\omega t}.$$

The Fourier coefficients are given by the formula

$$f_n(\omega) = \frac{\omega}{2\pi} \int_{-\pi/\omega}^{\pi/\omega} e^{-in\omega\tau} f(\tau) d\tau$$

The expansion in a Fourier series means that an arbitrary periodic function with a period $2\pi/\omega$ can be written in the form of a superposition (spectrum) of an infinitely large number of monochromatic functions with periods $2\pi/\omega$, $2\pi/2\omega$, ..., $2\pi/n\omega$ or frequencies ω , 2ω , ..., $n\omega$ and so on.

Conditions under which the expansion in a Fourier series is possible are usually fulfilled in physical applications.

Passing to the limit where the period increases indefinitely (i.e. $\omega \rightarrow 0$) and the frequencies draw nearer together, one can obtain the expansion in a Fourier integral:

$$f(t) = \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega. \quad (\text{II.1})$$

The function $F(\omega)$, called the Fourier transform of the function $f(t)$, is given by the formula

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt. \quad (\text{II.2})$$

The expansion in a Fourier integral is possible if the properties of $f(t)$ ensure the convergence of (II.1)–(II.2). In physical applications $f(t)$ usually tends to zero as $t \rightarrow \pm \infty$, which ensures the convergence of these expressions.

The Fourier integral can be written in a more symmetric form:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} F(\omega) d\omega, \quad (\text{II.3})$$

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt. \quad (\text{II.4})$$

If $f(t)$ is a real function, then $F(\omega)$ is a complex function, and

$$F^*(\omega) = F(-\omega). \quad (\text{II.5})$$

Formulae (II.3) and (II.4) can be combined in the form

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} d\omega \int_{-\infty}^{\infty} d\tau e^{-i\omega \tau} f(\tau), \quad (\text{II.6})$$

which is also often called the Fourier integral.

Formula (II.3) shows that $f(t)$ represents the sum of monochromatic terms $e^{i\omega t}$ which are taken with weights (amplitudes) $F(\omega) d\omega/\sqrt{2\pi}$

The complex amplitude $F(\omega)$ can be written in the form

$$F(\omega) = A(\omega) e^{i\varphi(\omega)}, \quad (\text{II.7})$$

where $A(\omega)$ is the modulus and $\varphi(\omega)$ is the phase of the function $F(\omega)$; they are real functions of the frequency ω . In such a representation for the Fourier integral we have

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} A(\omega) e^{i(\omega t + \varphi(\omega))} d\omega. \quad (\text{II.8})$$

We prove an important equality which is sometimes called the Parseval relation for the Fourier integral:

$$\int_{-\infty}^{\infty} (f(t))^2 dt = \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega. \quad (\text{II.9})$$

Now,

$$\begin{aligned} \int_{-\infty}^{\infty} (f(t))^2 dt &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \left\{ \int_{-\infty}^{\infty} A(\omega) e^{i(\omega t + \varphi(\omega))} d\omega \right\} dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} A(\omega) e^{i\varphi(\omega)} \left\{ \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt \right\} d\omega. \end{aligned}$$

But, by the definition (II.4),

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt = F^*(\omega) = A(\omega) e^{-i\varphi(\omega)}$$

Hence

$$\int_{-\infty}^{\infty} (f(t))^2 dt = \int_{-\infty}^{\infty} (A(\omega))^2 d\omega = \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega ,$$

which was to be proved.

The Delta-function and its Properties

The δ -function was introduced by Dirac * and proved to be very useful in considering many problems of theoretical physics.

The δ -function is defined by the relations

$$\delta(x) = 0 \quad \text{for} \quad x \neq 0 ,$$

$$\delta(x) = \infty \quad \text{for} \quad x = 0 ,$$

so that

$$\int_a^b \delta(x) dx = 1 , \quad \text{where } a < 0 < b . \quad (\text{III.1})$$

The basic property of the δ -function

$$\int_a^b f(x) \delta(x) dx = f(0) , \quad a < 0 < b , \quad (\text{III.2})$$

* P.A.M.Dirac, *The principles of quantum mechanics* (Clarendon Press, Oxford, 1958).

(where $f(x)$ is an arbitrary continuous function of x) follows immediately from the definition (III.1).

Indeed, because of the properties of the δ -function, only the neighbourhood of the point $x = 0$ is of importance in the integral (III.2). Then for the point $x = 0$ the function $f(x)$ can be brought out of the integral, and the remaining integral is equal to unity by virtue of (III.1). The integral (III.2) can also be rewritten in the form

$$\int f(x) \delta(x - x_0) dx = f(x_0). \quad (\text{III.3})$$

The range of integration in (III.3) must include the point $x = x_0$, otherwise the integral reduces to zero:

$$\int_a^b f(x) \delta(x - x_0) dx = 0, \quad \begin{cases} x_0 > b, \\ x_0 < a \end{cases} \quad (\text{III.3}')$$

The delta-function cannot be contained in any final expression. When the δ -function is written a subsequent integration with respect to the variables on which it depends is always implied; the δ -function can be considered as a limit of consistency for analytical functions.

In particular, such properties are possessed by the expression

$$F(\alpha, x) = \frac{\sin \alpha x}{\pi x},$$

which behaves as $\delta(x)$ when $\alpha \rightarrow \infty$. Indeed, for $x = 0$, $F(\alpha, x)|_{x=0}$ is equal to α/π and diverges as $\alpha \rightarrow \infty$. For $x \neq 0$, $F(\alpha, x)$ strongly oscillates about the zero value with a damped amplitude. Finally,

$$\int_{-\infty}^{\infty} \frac{\sin \alpha x}{\pi x} dx = 1$$

for any α . Consequently, we see that

$$\lim_{\alpha \rightarrow \infty} \frac{\sin \alpha x}{\pi x} = \delta(x). \quad (\text{III.4})$$

The integral of the form

$$\int_{-\infty}^{\infty} e^{ikx} dk,$$

which is often encountered and which should be understood as

$$\lim_{\alpha \rightarrow \infty} \int_{-\alpha}^{\alpha} e^{ikx} dk = \lim_{\alpha \rightarrow \infty} \frac{2}{x} \sin \alpha x$$

can be expressed in terms of the δ -function. Comparing with (III.4), we obtain

$$\int_{-\infty}^{\infty} e^{ikx} dk = 2\pi\delta(x). \quad (\text{III.5})$$

Other representations of the δ -function are

$$\delta(x) = \frac{1}{\pi} \lim_{\alpha \rightarrow 0} \frac{\alpha}{x^2 + \alpha^2}, \quad (\text{III.4}')$$

$$\delta(x) = \frac{1}{\sqrt{\pi}} \lim_{\alpha \rightarrow 0} \alpha e^{-x^2/\alpha^2}, \quad (\text{III.4}'')$$

$$\delta(x) = - \lim_{\alpha \rightarrow 0} \frac{e^{x/\alpha}}{\alpha(e^{x/\alpha} + 1)^2}. \quad (\text{III.4}''')$$

The delta-function can also be defined as the derivative of a particular discontinuous function $\epsilon(x)$:

$$\begin{aligned} \epsilon(x) &= 0, & x < 0, \\ \epsilon(x) &= 1, & x > 0. \end{aligned} \quad (\text{III.6})$$

It is obvious that $\epsilon'(x) = 0$ for $x \neq 0$. We show that the equality (III.2) also

holds:

$$\begin{aligned}\int_a^b f(x) \epsilon'(x) dx &= f(x) \epsilon(x) \Big|_a^b - \int_a^b \epsilon(x) f'(x) dx \\ &= f(b) - \int_0^b f'(x) dx = f(0) .\end{aligned}$$

Consequently,

$$\epsilon'(x) = \delta(x) . \quad (\text{III.7})$$

We list certain basic properties of the δ -function:

$$\begin{aligned}\delta(-x) &= \delta(x) , \\ \delta'(-x) &= -\delta'(x) , \\ x\delta(x) &= 0 , \\ x\delta'(x) &= -\delta(x) , \\ \delta(ax) &= \frac{1}{a} \delta(x) , \quad a > 0 , \\ \delta(x^2 - a^2) &= \frac{1}{2a} [\delta(x - a) + \delta(x + a)] , \\ \int \delta(a - x) \delta(x - b) dx &= \delta(a - b) , \\ f(x) \delta(x - a) &= f(a) \delta(x - a) , \\ \int f(x) \delta'(x - a) dx &= -f'(a) , \\ \delta(f) df &= \delta(x) dx , \\ \delta[f(x)] &= \frac{1}{|df/dx|} \delta(x - x_0) ,\end{aligned} \quad (\text{III.8})$$

where x_0 are the roots of the equation $f(x_0) = 0$,

$$\int \varphi(x) \delta[f(x) - \alpha] = \left\{ \frac{\varphi(x)}{|df/dx|} \right\}_{f(x)=\alpha},$$

$$\delta(\mathbf{r} - \mathbf{r}_0) = \delta(x - x_0) \delta(y - y_0) \delta(z - z_0), \quad (\text{III.8'})$$

$$\delta(r - vt) = \frac{1}{2\pi} \int e^{i(kr - \omega t)} \delta(\omega - kv) dk d\omega.$$

Since the δ -function has a meaning only provided that integration with respect to its argument is implied, then these equalities also mean that, multiplying the left-hand side of each of them by the continuous function $f(x)$ and integrating with respect to x , we shall arrive at the same results as those which will be given by the right-hand sides. Let us prove, for example, the third relation. For this we consider the integral

$$\int f(x) x \delta(x) dx = f(x) x|_{x=0} = 0$$

and the relation is proved.

The δ -function often proves to be useful in considering Fourier transforms. Thus, if we have the expansion of a certain function $f(x)$ in a Fourier transform

$$f(x) = \int_{-\infty}^{\infty} c(k) e^{ikx} dk, \quad (\text{III.9})$$

then, making use of (III.5), we immediately obtain the expression for the inverse Fourier transform. Indeed, multiplying the left-hand and right-hand sides of the equality (III.9) by $e^{-ik'x}$ and integrating with respect to x , we obtain

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) e^{-ik'x} dx &= \int_{-\infty}^{\infty} c(k) e^{i(k-k')x} dk dx = \\ &= \int_{-\infty}^{\infty} c(k) dk \int_{-\infty}^{\infty} e^{i(k-k')x} dx = \\ &= \int_{-\infty}^{\infty} c(k) 2\pi \delta(k-k') dk = 2\pi c(k'). \end{aligned}$$

Consequently, we have

$$c(k') = \frac{1}{2\pi} \int f(x) e^{-ik'x} dx. \quad (\text{III.10})$$

Analogous relations are also obtained in more complex cases. Formula (III.5) can be considered as the expansion of the δ -function in a Fourier transform.

We now consider the Legendre polynomial expansion of the δ -function:

$$\delta(1 - \xi) = \sum_l B_l P_l(\xi).$$

We find the coefficients B_l by multiplying the left-hand and right-hand sides of the equality by $P_l(\xi)$ and integrating with respect to ξ . Taking into account that

$$\int_{-1}^{+1} P_l(\xi) P_l(\xi) d\xi = \frac{2}{2l+1} \delta_{ll},$$

and

$$P_l(1) = 1,$$

we obtain

$$B_l = \frac{1}{2} (2l+1).$$

Finally we have:

$$\delta(1 - \xi) = \frac{1}{2} \sum_l (2l+1) P_l(\xi). \quad (\text{III.11})$$

Let us prove the following relation:

$$\nabla^2 \left(\frac{1}{r} \right) = -4\pi \delta(\mathbf{r}), \quad (\text{III.12})$$

where

$$\delta(\mathbf{r}) = \delta(x) \delta(y) \delta(z).$$

For this we expand the function $1/r$ in a three-dimensional Fourier integral:

$$\frac{1}{r} = \int c(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}} dk_x dk_y dk_z. \quad (\text{III.13})$$

Correspondingly, for the function $c(\mathbf{k}) \equiv c(k_x, k_y, k_z)$ we have, making use of (III.10),

$$c(\mathbf{k}) = \frac{1}{(2\pi)^3} \int \frac{1}{r} e^{-i\mathbf{k} \cdot \mathbf{r}} dV \quad (\text{III.14})$$

In formula (III.14) we integrate first with respect to the angle, choosing the polar axis in the direction of the vector \mathbf{k} :

$$\begin{aligned} c(\mathbf{k}) &= \frac{1}{(2\pi)^2} \int_0^\infty \frac{1}{r} r^2 dr \int_0^\pi e^{-ikr \cos \vartheta} \sin \vartheta d\vartheta \\ &= \frac{1}{(2\pi)^2} \frac{i}{k} \int_0^\infty (e^{-ikr} - e^{ikr}) dr. \end{aligned}$$

The last integral is usually calculated by multiplying the integrand by the factor $e^{-\alpha r}$ and by subsequently letting $\alpha \rightarrow 0$. We obtain finally

$$c(\mathbf{k}) = \frac{1}{(2\pi)^2} \frac{2}{k^2}.$$

Substituting $c(\mathbf{k})$ into (III.13), we have

$$\frac{1}{r} = \frac{2}{(2\pi)^2} \int \frac{1}{k^2} e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k}.$$

Taking the Laplacian of the left-hand and right-hand sides, we obtain

$$\nabla^2 \left(\frac{1}{r} \right) = - \frac{2}{(2\pi)^2} \int e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k},$$

and, consequently (see III.5)

$$\nabla^2 \left(\frac{1}{r} \right) = - \frac{2}{(2\pi)^2} (2\pi)^3 \delta(x) \delta(y) \delta(z) = -4\pi \delta(\mathbf{r}). \quad (\text{III.15})$$

In conclusion we note that, as follows from (III.1), the δ -function is a function with dimensions, its dimensionality being the inverse of that of its argument.

It is often convenient to make use of the sign function $\operatorname{sgn} x$ which is defined as

$$\operatorname{sgn} x = \begin{cases} +1 & x > 0, \\ -1 & x < 0 \end{cases} \quad (\text{III.16})$$

Its Fourier transform F_{sgn} and the integral representation is:

$$\begin{aligned} F_{\operatorname{sgn}}(\omega) &= \lim_{\alpha \rightarrow 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{sgn} x e^{-\alpha|x|} e^{i\omega x} dx = \\ &= \lim_{\alpha \rightarrow 0} \frac{i}{\pi} \frac{\omega}{\omega^2 + \alpha^2} = \frac{i}{\pi} P\left(\frac{1}{\omega}\right), \end{aligned} \quad (\text{III.17})$$

where P denotes the principal value of the function $1/\omega$:

$$P(1/x) = \begin{cases} 1/x & x \neq 0, \\ 0 & x = 0. \end{cases} \quad (\text{III.18})$$

In this case

$$\int f(x) P\left(\frac{1}{x}\right) dx = P \int \frac{f(x)}{x} dx, \quad (\text{III.19})$$

where $P \int$ is the principal value of the integral.

The inversion gives

$$\operatorname{sgn} x = \frac{i}{\pi} P \int_{-\infty}^{\infty} \frac{e^{-i\omega x}}{\omega} d\omega. \quad (\text{III.20})$$

SUBJECT INDEX

- Absorption of radiation, 163
- Accelerated motion, 241
 - in relativistic mechanics, 335
- Acceleration, 260
 - , four, 252
 - of charge, 351
 - , self-, 123
- Action at a distance, 93, 237
- Addition of velocities, Einstein's law of
 - 235, 248
- Advanced potential, 92, 100
 - , boundary condition for, 93
- Aether, universal, 220
- Algebra, vector, 356
- Amplitude of plane wave, 151
- Analysis, vector, 3, 356
- Angle, transformation of an element of
 - solid, 316
- Angular momentum, 84
 - of interacting particles, 349
 - of the electromagnetic field, 48
 - transformation, 310
 - , relativistic, 236
- Annihilation, positron-electron, 284
- Antisymmetric tensor, 254
- Atomic model, planetary, 196
 - nuclei, quadrupole moment, 62
 - , stability of, 276
- Axial vector, 23, 358

- Balance of forces, 121
- Betatron, 338
 - mode, 339
- Binding energy, nuclear, 277
- Biot-Savart law, 75
- Boundary condition, 164
 - for advanced potential, 93
 - for electromagnetic potentials, 86
 - for retarded potential, 93
 - for scalar potential, 51
 - for vector potential, 95
- Bremsstrahlung, 209, 353, 355

- Carrier frequency, 152
- Causal description, 94
- Centre of charge, 61
- Centre of mass, 348
- Centre-of-mass system, 190, 272
 - , cross-section in the, 208
 - , laboratory system and, 208
 - , velocity of, 273, 349
- Charge (*see also* charged particle), 10, 13
 - , centre of, 61
 - conservation law, 11, 17, 29, 87, 292
 - , continuous distribution of, 14
 - density, 14, 52, 109
 - and Lorentz transformation, 293
 - dipole interaction, 65
 - , sign of the, 10
 - , total, 109
- Charged particle (*see also* charge)
 - , accelerated, 117
 - , electric field of uniformly moving
 - relativistic, 298
 - , electromagnetic field of, 135
 - , electromagnetic field of harmonic
 - moving, 118, 119, 125, 137
 - , electromagnetic field of relativistic,
 - 296
 - , equations of motion for, 328, 329,
 - 341
 - in electric field, 174, 332, 334
 - in electromagnetic field, 11, 179,
 - 187, 327
 - in magnetic field, 118, 176, 182, 184

- Charged particle (continued)
 - , magnetic field of uniformly moving relativistic, 299
 - , radiation of moving, 106
 - , radiation of relativistic, 351
 - , system of two, 120
 - , two, interaction force between, 189, 300
 - sphere, uniform motion of, 301
- Circular polarization, 152
- Circular of vector, 3, 360
- Classical field theory, 11
 - invariance transformation, 243
- Coherent scattering, 157
- Collision, 197
 - between relativistic particles, 270, 286
 - , elastic, 198
 - , electron-nucleon, 353
 - , head-on, 287
 - of particles with nonzero rest mass, 286
 - , photon-electron, 289
 - process, irradiated energy in, 212
 - , proton-proton, 285
 - , relativistic particles in, 270
- Compton effect, 289
 - scattering, 289
 - wavelength, 124, 290
- Conservation, charge, 11, 87, 292
 - law, charge, 17, 29
 - , energy, 42, 44, 265, 279
 - , energy-momentum, 271
 - for the motion in a constant magnetic field, 178
 - , momentum, 45, 265, 283
- Continuity, equation of, 18
- Contraction of a moving scale, 233
- Converging spherical wave, 90
- Coordinates, spherical system of, 114
- Coulomb gauge, 40
 - law, 10, 53
 - scattering, differential cross section for 203
- Cross section, differential, 158, 197
 - , differential, for Coulomb scattering, 203
 - in the centre-of-mass system, 208
 - in the laboratory system, 208
 - , scattering, for unpolarized radiation, 159
 - , total, 198
 - , total scattering, transformation of the, 319
- Curl of a vector, 365
- Current density for quasistationary motion, 70
 - , scalar potential and, 110
 - vector, 16
 - , vector potential and, 110
 - , displacement, 28, 30
 - , character of displacement, 33
 - , electric, 16
 - , four-dimensional, 292
 - , harmonic, 100
 - , magnetic moment and closed, 83
 - , non-steady, 23
 - , total, 28
- Curvilinear coordinates, 370
- Cyclotron, 338
 - frequency, 117, 337
- Cylindrical polar coordinates, 372
- D'Alembertian, 36
- D'Alembert's equation, 36, 94
 - method, 88
 - operator, 294
- Damped oscillator, 126
- Damping coefficient, 126
- Decay, meson, 279
 - , μ -meson (muon), 281
 - , particle, 275
 - , elementary, 279
 - , spontaneous, 276
 - , π^{\pm} -meson, 280
 - , π^0 -meson, 282
- Deceleration of fast electrons in matter, 355
- Defect, mass, 277
- Deflection angle, 199
 - of particles in a magnetic field, 338
- Delay time, 91
 - , proper, 108
- Delta-function, 379
 - , basic properties of, 379, 382
 - , Legendre polynomial expansion of, 384

- Differential cross section, 158, 197
- Dimensions of wave packet, 154
- Dipole, 55
 - approximation, 55, 111
 - of retarded potential, 106
 - dipole interaction, 65
 - electric field, 57
 - moment, 55
 - and the origin, 56
 - , changing, 112
 - density, 56
 - , orientation of, 65
 - radiation, 116, 341
 - , magnetic, 131
- Dispersion formula of classical electrodynamics, 160
- Displacement current, 28, 30, 73
 - , character of, 33
 - , experimental verification of, 31
 - , introduction of, 32
 - , logical necessity of, 31
- Divergence of a vector, 364
- Diverging spherical wave, 90
- Doppler effect, 309
 - , longitudinal, 310
 - , second order, 312
 - , transverse, 311
- Drift, 180
- Dyadic product of vectors, 369
- Dynamics, equations of, 256

- Effective potential energy, 192
- Einstein's formulae, 264
 - law of addition of velocities, 235, 248
- Electric and magnetic field, symmetry between, 33
 - field, 12, 112
 - and slow uniform motion, 77
 - , determination of, 26
 - , dipole, 57
 - of a system of point charges, 52
 - of moving charge, 298
 - , particle accelerated by an, 334
 - , quadrupole, 61
 - , relativistic, 304
 - strength, 12
 - , transverse, trajectory of charge in, 333
 - , uniform, charged particle in, 174
 - , uniform, relativistic particle in, 331
- Electrodynamics, dispersion formula of classical, 160
- Electromagnetic field, 18
 - as a mechanical system, 170
 - , charged particle in, 179, 187
 - energy, 42, 67
 - equation, 164
 - in induction zone, 140
 - in radiation zone, 140
 - invariances, 307
 - of moving charge, 296
 - , particle trajectory in, 181
 - , reality of the, 43
 - strength in the plane wave, 146
 - tensor, 305
 - induction, 23
 - interaction, 10
 - potential, boundary conditions for, 86
 - , equations for, 85
 - , initial conditions for, 86
 - wave, propagating, 143
 - , scattering of the, 157
- Electromotive force, 16
- Electron, collision, photon-, 289
 - fast, deceleration of, in matter, 355
 - nucleon collision, 353
 - positron pair production, 283
 - radius of, classical, 266
- Electrostatic field, 12
 - , Maxwell's equations for the, 15
 - , potential energy of the, 63
 - , work performed by the, 13
 - potential, 16
- Electrostatics, direct problem of, 50
 - equations of, 15
 - inverse problem of, 50
- Elementary particle, 9
 - decay, 279
- Elliptical polarization, 152
- E.m.f., 16
- Emission of energy, rate of, 352
- Emitted energy, flux density of, 115
 - , spectral distribution of, 137
 - , total flux of, 116
 - line, half-width of, 128

- Emitter, 93
 - of radiation, 112
- Emitting particle, momentum loss of, 117
 - system, Poynting vector of, 115
- Energy balance, 122
 - conservation law, 42, 44, 265, 279
 - , electromagnetic field, 42
 - , equivalence of mass and, 264
 - , field, 171
 - flux, 43
 - - of radiation, 117
 - , interaction, 341
 - interaction, of system of charges, 65
 - , kinetic, 265
 - momentum conservation law, 271
 - - four-vector, 262
 - of a particle, 260, 263
 - of interacting particles, 347
 - of nuclear reaction, 278
 - , radiation, 350
 - , rest, 264
 - , total, 265
 - , transformation of an electromagnetic wave, 317
- Equipotential surface, 361
- Equivalence of mass and energy, 264
- Ether theory, 31
- Experimental verification of displacement current, 31
 - - of Lorentz contradiction, 321
- Faraday law of induction, 35
 - - of induction (integral form), 24, 26
- Field, electric, 12
 - , electric, strength, 12
 - energy, electromagnetic, 42.
 - lines, 362
 - oscillators, 167
 - - , energy of, 173
 - - , number of, 171, 173
 - , scalar, 358
 - , self, 29
 - , vector, 3, 362
- Fluorescence, resonance, 160
- Flux density, 197
 - - of emitted energy, 115
 - , total, of emitted energy, 116
 - , vector, 3, 362
- Focussing, 178
- Force on a system of charges, electric, 64
 - , transformation of a, 326
- Four acceleration, 252
 - dimensional current, 292
 - - wave vector, 308
 - force, 257
 - - , fourth component of, 258
 - momentum, 257, 262
 - potential, 294
 - - , Lienard-Wiechert, 302
 - tensor, 253
 - - , invariants of, 255
 - - , second-rank, 253
 - - , transformation of, 254
 - vector, 250
 - - , energy-momentum, 262
 - velocity, 251
- Fourier coefficients, 375
 - integral, 375, 376
 - series, 375
 - transform, 376
 - transformation of vector potential, 96
- Fourth coordinate, 244
- Frequencies, transformation of, 308
- Galilean principle of relativity, 219
 - transformation, 218
 - - , Lorentz and, 227
- Gauge condition, 295
 - , Coulomb, 40
 - invariance, 38
 - , Lorentz, 39
 - of the potentials, 38
 - transformation, 38, 147
- Gauss-Ostrogradsky theorem, 4, 14, 363
- Generalized force, 329
 - momentum, 329
- Gradient of scalar field, 358
- Group velocity, 154
- Half-width of emitted line, 128
- Hamiltonian of a particle in an electromagnetic field, 330
 - operator, 359, 362, 367, 372
- Hamilton's equations for field oscillators, 167
- Helicity, 152

- Homogeneity of space, 225
- Imaginary time, 244
- Impact parameter, 199
 - and scattering angle, 201
- Induction accelerator, 338
 - , Faraday law of, 35
 - , Faraday law of (integral form), 24, 26
 - zone, 139
- Inertial reference frame, 217
- Initial conditions for electromagnetic potentials, 86
- Interacting particles, angular momentum of, 349
 - , energy of, 347
 - , Lagrangian for, 346
 - , mass of, 348
 - , momentum of, 348
 - , potential of, 342
 - , system of two, 350
- Interaction, electromagnetic, 10
 - energy, generalized, 346
 - force between two charges, 300
 - , interparticle, relativistic, 269, 274
 - of charges, 11
 - , particles in electromagnetic, 340
 - potential, experimental determination of, 203
 - , principle of the limiting velocity of propagation of, 223
 - , retarded, 340
- Interval between events, 239
 - , space-like, 240
 - , time-like, 240
- Invariance, electromagnetic field, 307
 - in Minkowski space, 245
 - of physical laws under a rotation, 242
 - of the interval, 239
 - , phase, 307
 - , relativistic, 222
- Invariant of four-tensor, 255
 - physical law, 217
- Ion, 11
- Irradiated energy in collision process, 212
- Irrotational vector, 360, 366
- Ives' and Stillwell's experiment, 311
- Laboratory system, 206
 - and centre-of-mass system, 208
 - , cross section in the, 208
- Lagrange equation for charged particle, 328, 329
 - for charged particle in electromagnetic field, 187
 - , relativistic, 267
- Lagrangian for interacting particles, 346
 - of a particle in an electromagnetic field, 329
 - of a system of charges, 341
 - , relativistic, 267
- Lamé coefficients, 371
- Laplacian, 5, 374
- Larmor frequency, 177
- Law of motion, 217
- Legendre polynomial expansion of the delta-function, 385
- Lenz rule, 25
- Lienard-Wiechert four-potential, 105, 302, 350
- Light pressure, 47
 - , velocity of, 20
- Line-width, radiation reaction and, 125
- Linear radiator, 140
- Long-range action, 13
- Longitudinal Doppler effect, 310
- Lorentz and Galilean transformation, 227
 - condition, 93, 111
 - contraction, 229
 - , experimental investigation of, 321
 - force, 20, 29, 121, 326
 - in terms of electromagnetic potential, 328
 - friction force, 126
 - gauge, 39
 - invariant relations, 242
 - relation, 36
 - spectral distribution function, 126
 - transformation, 226
 - , charge density and, 293
 - , graphical representation of, 247
 - of potentials, 295
- Magnetic dipole radiation, 131, 132

- Magnetic field, 20, 112
 - and quasistationary motion, 75, 81
 - and slow uniform motion, 78
 - , deflection of particles in a, 338
 - , determination of a, 22, 26
 - , nonuniform, charged particle in, 184
 - of moving charge, 299
 - , relativistic, 305
 - strength, 20
 - , symmetry between electric and, 33
 - , time dependent, 24
 - , trajectory of charge in, 337
 - , uniform, charged particle in, 176
 - , varying in time, uniform, charged particle in, 182
 - vector lines, 21
- Magnetic moment, 81
 - and the origin, 82
 - and closed current, 83
 - , conservation of, 183, 185
- Magnetomotive force, 20, 25
- Mass and energy, equivalence of, 264
 - defect, 277
 - energy relation, 284
 - of interacting particles, 348
 - , relativistic, 261
 - , conservation, 265
 - , rest, 256, 261
 - , zero, 261
- Material point, 256
- Maxwell equations, 28
 - for static problems, 49
 - for the electrostatic field, 15
 - , quasistationary, 71, 73
 - Lorentz equations, 30, 294, 305
- Meson decay, 279
- Michelson's experiment, 221, 230
- Minkowski force, 257
 - space, 244
 - , invariance in, 245
 - , rotation transformation in, 246
- Microparticles, 9
- Microsystems, 9
- Mirror, magnetic, 186
- Momentum conservation law, 45, 265, 283
 - density of the electromagnetic field, 46
 - , generalized, 189
 - loss of emitting particle, 117
 - of a plane wave, 144
 - of interacting particles, 348
- Mössbauer effect, 312
- Motion for charged particle, equations of, 328
- Moving clock, slowing down of the running of, 231
 - particle, radiation emitted by a, 315
 - scale, contraction of a, 233
- Multiplication, vector, 356
- Mu-meson (μ on) decay, 232, 281
 - , energy, 281
- Neutral particle, 10
 - system, electrically, 55, 111
- Neutrino, 280
- Newtonian law of inertia, 256
- Normalization cubes, 164
- Nuclear binding energy, 277
 - reaction, energy of, 278
- Nucleon collision, electron-, 353
- Oersted, law of, 21
- Orthogonality, condition of, 249
- Parseval relation, 377
- Period of motion, 71
- Phase invariance, 307
 - of plane wave, 151
 - shift, 158
 - velocity, 144, 154
- Photon-electron collision, 289
- "Picture", 322
- Pi^{\pm} -meson decay, 280
- Pi^0 -meson decay, 282
- Plane polarization, 152
- Plane wave, amplitude of, 151
 - , electromagnetic field strength in the, 146
 - , momentum of a, 144
 - , phase of, 151
 - , Poynting vector in a, 144

- Point charges, electric field of a system
 of, 52
 --, moving, 102
 --, system of, 52
 Poisson's equation, 5, 50, 74
 Polar coordinates, cylindrical, 372
 --, spherical, 372
 -- vector, 23, 358
 Polarization, circular, 152
 -- direction, 152
 --, elliptical, 152
 --, plane, 152
 Positron-electron annihilation, 284
 Potential, 360
 -- energy, effective, 192
 -- of electrostatic field, 63
 -- of system of charges, 63
 --, Lorentz transformation of, 295
 -- of interacting particles, 342
 --, retarded, 91
 -- vector, 360
 --, --, 7, 34, 91, 131
 Poynting vector, 43
 -- in a plane wave, 144
 -- of emitting system, 115
 Principal value, 386
 Production, pair, electron-positron, 283
 Propagating electromagnetic wave, 143
 Propagation of interaction, principle of
 the limiting velocity of, 223
 Proper delay time, 108
 -- retardation, 107, 138
 -- time, 231, 241
 Proton-proton collision, 285
 Pseudo-vector, 23, 358
- Quadrupole, 58
 -- electric field, 61
 -- moment, 58
 -- and symmetry, 62
 -- density, 58
 -- of atomic nuclei, 62
 --, sign of the, 61
 -- tensor, 60
 -- radiation, 129, 131
 --, intensity of, 132
- Quasistationary condition, 72
 -- Maxwell equations, 71, 73
 -- motion, 69
 --, current density for, 70
 --, magnetic field and, 75, 81
 --, vector potential and, 74, 80
- Radiated power, total, 116, 117
 --, transformation of, 317
 Radiation, dipole, 116
 -- emission, 44, 112
 -- emitted by a moving particle, 315
 -- energy, 350
 -- and momentum, 116
 --, energy flux of, 117
 -- from a system of two charges, 120, 195
 -- of a beam of particles, effective, 354
 -- of moving charge, 351
 -- power of oscillating charge, 119
 -- pressure, 47
 -- reaction, 121, 124
 -- and line-width, 125
 --, order of magnitude of, 123
 -- recording, 322
 --, scattered, 158
 --, scattering cross section for unpolarized
 159
 -- zone, 114, 139
 Radioactivity, 275
 Radius, classical, of electron, 124, 266
 -- vector, four-dimensional, 249
 Rayleigh scattering, 161
 Recoil shift, 312
 Reduced mass, 84, 191
 Reference frame, 216
 --, inertial, 217
 Relative motion of interacting particles,
 192
 -- velocity, 84
 Relativistic angular transformations, 236
 -- dynamics, equations of, 257
 -- electric field, 304
 -- invariance, 222, 244
 -- Lagrange equations, 267
 -- magnetic field, 305
 -- mass, 261

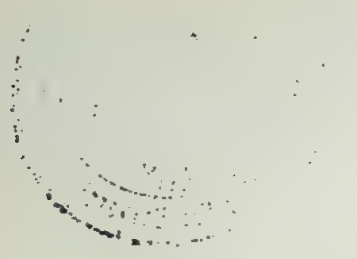
- Relativistic mass, conservation of, 265
 - particles, collisions between, 270, 286
 - in uniform electric field, 331
 - system of particles, 269
- Relativistically invariant relations, 242
- Relativity of spatial extension, 228
 - , special theory of, postulates of, 222
- Resonance fluorescence, 160
- Rest energy, 264
 - mass, 256, 261, 273
 - of a system of particles, 274
 - , zero, 261
- Retardation, proper, 107
 - , scalar potential and, 344
 - , total, 342
 - , vector potential and, 345
- Retarded argument, 113
 - interaction, 340
 - potential, 91, 100, 102
 - , boundary condition for, 93
 - , dipole approximation of, 106
- Rigid body, 230
- Rotation, invariance of physical laws under, 242
 - transformation in Minkowski space, 246
- Rutherford formula, 203
- Scalar field, 358
 - potential, 5, 50, 90
 - and current density, 110
 - and retardation, 344
 - , boundary condition for, 51
 - , determination of, 37
 - , equation for, 35
 - product, 250
- Scale, 228
- Scattered particle, energy lost by, 209
 - radiation, 158
- Scattering angle, 199
 - , impact parameter and, 201
 - , coherent, 157
 - , Compton, 289
 - cross section for unpolarized radiation, 159
 - of the electromagnetic wave, 157
 - process, 197
- Self energy of a point charge, 68
- Separation constant, 166
- Shift, frequency, due to thermal motion, 313
- Short-range action, 13
- Sign function, 386
- Signal, 224
 - , propagation of, 224
 - , transmission, 236
- Simultaneity, 237
 - , absolute, 237
- Sinks, 364
- Slow uniform motion, electric and magnetic field, 77, 78
- Solenoidal vector, 366
 - field, 6, 364
- Sources, 364
- Space-like interval, 240
- Space-like vector, 251
- Spatial extension, relativity of, 228
- Special theory of relativity, postulates of, 222
- Spectral decomposition, 135
 - distribution functions, Lorentz, 128
 - of emitted energy, 137
- Spherical polar coordinates, 372
 - system of coordinates, 114
 - wave, 114
 - , converging, 90
 - , diverging, 90
- Spontaneous particle decay, 276
- Stability of atomic nuclei, 276
- Static problems, Maxwell's equations for, 49
- Stokes' theorem, 4, 366
- Superposition, principle of, 87
 - of electromagnetic fields, 29
 - property, 12
- Surface integral, 361, 363
- Symmetry between electric and magnetic field, 33
- Synchrotron, 338
- System of particles, rest mass of a, 274
- Target, 197
- Tensor, antisymmetric, purely, 254
 - , field, electromagnetic, 305

- Tensor (continued)
 - , four-, 253
 - , second-rank four-, 253
 - , transformation of four-, 254
- Test charge, 12
 - , moving, 20
 - particle motion, 174
- Thermal motion, frequency shift due to, 313
- Thomson atomic model, 119
 - scattering, 161
- Threshold kinetic energy, 283
- Time, imaginary, 244
 - -like interval, 240
 - -like vector, 251
 - , proper, 231, 241
 - reversal invariance, 92
- Trajectory, 217
 - , charge in magnetic field, 337
 - , charge in transverse electric field, 333
 - of interacting particles, 194
 - , particle in electric field, 175
 - , particle in electromagnetic field, 181
 - , particle in magnetic field, 177
- Transformation, invariance, classical, 243
 - of a force, 326
 - of an electromagnetic wave energy, 317
 - of an element of solid angle, 316
 - of radiated power, 317
 - of the total scattering cross section, 319
- Transverse Doppler effect, 311
 - waves, 166
- Trochoid, 181
- Ultrarelativistic motion, 268
- Uncertainty relations, 155
- Vector algebra, 356
 - analysis, 3, 356
 - , axial, 23, 358
 - , circulation, 3, 360
 - , curl of a, 365
 - field, 3, 362
 - , determination of a, 4, 8
 - , solenoidal, 6, 364
 - , vortex-free part of a, 6
 - flux, 3, 362
 - , irrotational, 360, 366
 - lines, 12
 - multiplication, 356
 - , polar, 23, 358
 - potential, 7, 34, 91, 131
 - , —, 360
 - and current density, 110
 - and quasistationary motion, 74, 80
 - and retardation, 345
 - , boundary condition for, 95
 - , condition on the, 8
 - , determination of, 37
 - , equation for, 35, 94
 - , Fourier analysis of, 170
 - , Fourier transformation of the, 96
 - of constant uniform field, 76
 - , pseudo-, 23, 358
 - , solenoidal, 366
 - , space-like, 251
 - , time-like, 251
- Velocity, four, 251
 - of centre of mass, 273, 349
- Vortex-free part of a vector field, 6
- Wave equation, 37, 88
 - number, 135
 - packet, 152
 - , dimensions of, 154
 - process, 90
 - vector, 135, 143
 - , four-dimensional, 308
- Wavelength, Compton, 124, 290



Theoretical Physics

Volume 2



Translated from the Russian by
S. Subotić, Belgrade

Translation edited by
J. Schneps, Tufts University, Medford, Mass.
A. J. Manuel, Leeds University



184461

QC51.2

'L 578

v.2



Theoretical Physics

An Advanced Text

Volume 2:

STATISTICAL PHYSICS

ELECTROMAGNETIC PROCESSES IN MATTER

Benjamin G. Levich

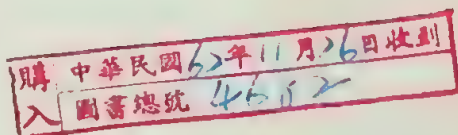
Institute of Electrochemistry

Academy of Sciences of the USSR, Moscow

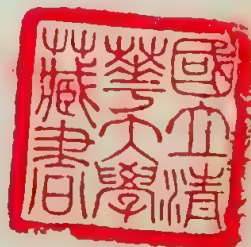


1971

North-Holland Publishing Company - Amsterdam • London



142686



© NORTH-HOLLAND PUBLISHING COMPANY, AMSTERDAM, 1971

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior permission of the Copyright owner.

Library of Congress Catalog Card Number: 68 54501

ISBN complete set: 0 7204 0176 3

Vol. 2: 0 7204 0178 X

Printed in The Netherlands

Title of the Russian edition:

KURS TEORETICHESKOJ FIZIKI

Russian edition published by:

IZDATELSTVO 'NAUKA', GLAVNAJA REDAKCIJA,

FIZIKO-MATEMATIČESKOJ LITERATURY (MOSKVA, 1969)

Publishers:

NORTH-HOLLAND PUBLISHING COMPANY - AMSTERDAM

Sole Distributors for the Western Hemisphere:

WILEY INTERSCIENCE DIVISION

JOHN WILEY & SONS, INC. - NEW YORK

FOREWORD

The first Russian edition of 'Theoretical Physics', which appeared in 1962, has been widely used as a textbook.

Numerous comments from colleagues, lecturers and students have been taken into account in preparing this new edition, which is the first one in English and which will also appear as the second Russian edition.

The material has now been divided into 4 volumes covering the following subjects

Volume 1

- Part I Theory of the Electromagnetic Field
- Part II Theory of Relativity

Volume 2

- Part III Statistical Physics
- Part IV Electromagnetic Processes in Matter

Volume 3

- Part V Quantum Mechanics

Volume 4

- Part VI Quantum Statistics and Physical Kinetics

The rapid development of physics and the present wide interest in non-equilibrium and non-stationary processes has compelled us to expand the section on physical kinetics. It has also been transferred to the end of Volume 4 as it is practically impossible to expound this topic without using quantum mechanics.

Part IV – 'Electromagnetic Processes in Matter' – has been substantially revised. Interest in this field has increased recently, mainly in connection with the study of plasmas and plasma-like media, which now have sections devoted to them.

The methods of calculating electrostatic and direct-current fields, and other problems of classical electrodynamics in a medium, are covered very briefly as we have assumed that students will be able to consult the many monographs and handbooks on general physics, electrical- and radio-technology, and the equations of mathematical physics.

As for other modifications and additions, we should draw attention to the introduction of tensor notation, to new ideas in the theories of relativity and electromagnetic fields, the broadening of the introduction to the theory of probability, a brief presentation of the method of correlation functions in statistical physics, the exposition of the thermodynamic theory of ferromagnetism and the theory of propagation of electromagnetic waves in plasma. A number of paragraphs have been rewritten. We have tried to bring the content of the book even closer to the interests of present-day theoretical physics.

The general level of the book has been preserved and it is still intended to form an introduction to theoretical physics. Problems requiring the use of cumbersome or special mathematical apparatus are still excluded, and the most difficult sections are marked by an asterisk. These may be skipped at will, since there is no reference to them in the main text.

In conclusion we would like to express our gratitude to all those who helped us in preparing this book, in particular to A.M. Brodsky, A.M. Golovin, B.M. Grafov, R.R. Dogonadze, V.S. Krylov and especially V.S. Markin and V.V. Tolmachev. I.V. Savelyev discovered a number of misprints which have now been corrected.

L.D. Konkina helped us in editing the manuscript.

We are grateful to the readers and students who used the first Russian edition of the book for sending us their valuable comments which have been taken into account in this edition.

August 1970

FOREWORD TO THE FIRST RUSSIAN EDITION

The continuous development of theoretical physics and the regular expansion of its areas of application create increasing demand for textbooks and manuals.

The rapid development and the complexity of the most recent experimental methods of physical investigation, and the corresponding development and extension of the mathematical apparatus of theoretical physics, have meant that one man usually cannot combine the two methods of investigation. The end of the 19th century and particularly the 20th century therefore saw physicists divided into 'experimentalists' and 'theoreticians', the latter studying physical laws by means of the mathematical methods of theoretical physics.

Obviously, a background in theoretical physics is essential in the education of experimental as well as theoretical physicists.

The experimental and theoretical methods of physical investigation have penetrated into a number of branches of science related to physics (physical chemistry, biophysics, geophysics, astrophysics, and so on) and into technology (metal physics and metallurgical science, thermophysics, electrical technology, radiotechnology, computation, the instrument-making industry etc.). Workers in these branches of science and technology also need a certain minimum knowledge of theoretical physics.

The compilation of a modern textbook on theoretical physics is inevitably associated with certain logical and methodological difficulties. It is impossible at present to divide theoretical physics into classical and quantum parts so that it is also impossible to divide it into separate chapters and sections. For example, the exposition of statistical physics without taking into account the quantum properties of atomic systems is impossible, for it would mean that the general theory remained without practical application. In the theory of electromagnetic processes in matter one has of necessity to make use of the ideas of statistical physics, and so on. It may be that the maximum consistency of composition would be obtained if the book were founded on

quantum mechanics but this is completely inadmissible in a book intended as an introductory treatise. Quantum mechanics requires a certain preparedness and the student must be convinced of the necessity of renouncing obvious classical representations. Compromise solutions, which have justified themselves during many years of teaching theoretical physics at the Moscow Engineering-Physical Institute and Moscow State University, are therefore inevitable.

The following general principles have been applied.

(1) The book is written as an introduction to theoretical physics so that aspects requiring the use of cumbersome or special mathematical apparatus have not been included.

(2) As it is to be used for a systematic study of the subject the course is a unique whole and all material necessary for understanding the later sections is contained in the earlier ones.

(3) It would not be feasible to elucidate experimental facts in addition to problems concerning purely theoretical physics. However, physics is a single science, and an attempt to expound the theoretical aspects without taking experiment into account would be quite wrong. The reader is assumed to have some basic experimental knowledge from university courses in general and atomic physics so that we have confined ourselves to references and, in a few instances, to a schematic description of basic experiments.

(4) The acquaintance assumed with general courses in general and atomic physics has allowed us to rely on a certain (very restricted) knowledge of quantum mechanics in our treatment of statistical physics.

(5) Classical mechanics usually forms a separate course so that this topic has been omitted although detailed reference has been made to handbooks of mechanics.

(6) The book similarly does not cover hydrodynamics, aerodynamics, the theory of heat transfer, or problems related to electrical- and radio-technology.

(7) Detailed reference is made to mathematical manuals. The mathematical apparatus utilized, except in the sections marked by an asterisk, is covered by the usual courses in analysis. In the case of quantum mechanics, however, the mathematical apparatus has been included, since it is of a specific character and is not taught in traditional mathematical courses.

(8) As the book is intended as a systematic course in theoretical physics no attempt has been made to achieve the same level of accessibility in all sections. It is a well-known fact that a student's comprehension and assimilation of difficult material increases as a course progresses, and that this is also true for the associated mathematical apparatus. Moreover, experi-

mental physicists will constantly encounter new problems in quantum mechanics which can only be handled using advanced methods of treatment. The section on quantum mechanics (Part V) therefore deals with some topics having a more advanced character than those in other sections. The analysis of applications of the kinetic equations is similarly treated rather extensively.

The uniqueness of the book's objectives has affected the content of individual sections, so that some topics in modern physics have been included at the expense of more traditional material.

Part I contains the foundations of the theory of the electromagnetic field in a vacuum, based on the system of Maxwell-Lorentz equations. A basic knowledge of electromagnetism is assumed. The focus of attention is the theory of radiation and the motion of charged particles in external fields.

In Part II, devoted to the theory of relativity, a four-dimensional form of representation is adopted which not only corresponds to the spirit of the theory but also predominates in contemporary literature. The problems of dynamics in the theory of relativity are treated in some detail. A number of the most recent applications of the theory of relativity, particularly those related to nuclear physics, are covered here for the first time in a textbook.

Part III is a revised version of Levich's 'Introduction to Statistical Physics' and treats statistical physics and the fundamentals of statistical thermodynamics. Classical thermodynamics would require too much space, and did not seem indispensable.

Part IV contains the theory of electromagnetic processes in matter. Relatively little attention is paid to problems in theoretical electrical- and radio-technology. The phenomenological theory of electric and magnetic properties of matter is analyzed in some detail, and the notion of the physics of the plasma state of matter is given.

In Part V the basic ideas of present-day relativistic quantum mechanics are included as well as the traditional problems of non-relativistic quantum mechanics. Applications to solid-state theory are considered at length.

Part VI contains the essential concepts of physical kinetics, which are not usually presented in a general course on theoretical physics.

The experience of teaching theoretical physics shows that the greatest difficulties are often encountered not in understanding new physical ideas but in the actual mathematical treatments. All mathematical operations have therefore been performed in sufficient detail.

For convenience we have presented a brief derivation of those formulae of

vector analysis which are encountered throughout, as well as the necessary data on Fourier integrals and δ -function theory.

The numbering of formulae and sections starts afresh in each Part and references to appendices have been given Roman numerals.

The author hopes that the readers, after making themselves familiar with the foundations of theoretical physics expounded in this book, will be able to proceed to a more profound study using the many-volume treatise of Landau and Lifshitz. The scientific and educational ideas of their work were of great influence on the author, who is a disciple of Landau.

Parts I–IV and Part VI were written by B.G. Levich. Part V was written by Y.A. Vdovin and V.A. Myamlin under the general scientific guidance of B.G. Levich. Chapter XV * of Part V was written by A.I. Naumov.

The author expresses his gratitude to the colleagues who read the book and the manuscripts, and made a number of valuable remarks: B.M. Grafov, R.R. Dogonadze, V.A. Kiryanov, V.S. Krylov, V.S. Markin, V.P. Smilga, Y.A. Chizmadzhev and Y.I. Yalamov.

The creation of a textbook on theoretical physics sufficiently comprehensive in content and clear in presentation is a very complex task. The author is therefore conscious of the fact that shortcomings and errors will be discovered and would be grateful to receive an account of them which can be taken into consideration in the next edition of the book.

1962

* Chapter 13 of the English edition.

Theoretical Physics: Outline of Vols. 1—4

Volume 1

Part I Theory of the Electromagnetic Field

- Chapter 1 General theory of the electromagnetic field
- 2 The electrostatic field
 - 3 The quasistationary magnetic field
 - 4 The electromagnetic field of arbitrarily moving charges
 - 5 Radiation theory
 - 6 Electromagnetic field in a vacuum and electromagnetic wave scattering
 - 7 The motion of particles in electromagnetic fields

Part II Theory of Relativity

- Chapter 1 General principles of the theory of relativity
- 2 Relativistic mechanics
 - 3 Relativistic electrodynamics

Appendix I, II and III

Subject index

Volume 2 (for details see p. xv)

Part III Statistical Physics

- Chapter 1 The basic concepts of the theory of probability
- 2 The kinetic theory of gases

- 3 Statistical distribution
- 4 Statistical and phenomenological thermodynamics
- 5 Ideal gases
- 6 Systems of interacting particles
- 7 Crystals
- 8 The theory of fluctuations
- 9 Systems with a variable number of particles
- 10 Statistical distributions in quantum statistics and some of their applications

Part IV Electromagnetic Processes in Matter

- Chapter 1 Electromagnetic fields in matter
- 2 Electrostatics
 - 3 Direct electric current and the magnetic properties of matter
 - 4 Quasistationary electromagnetic fields
 - 5 High-frequency fields
 - 6 Matter in the plasma state

Appendix IV

Subject index

Volume 3

Part V Quantum Mechanics

- Chapter 1 The basic concepts of quantum mechanics
- 2 The Schrödinger equation
 - 3 The mathematical apparatus of quantum mechanics
 - 4 Motion in a centrally symmetric field
 - 5 The quasi-classical approximation
 - 6 The matrix form of quantum mechanics
 - 7 Perturbation theory
 - 8 Spin and identity of particles
 - 9 Applications of quantum mechanics to the consideration of the properties of atomic and nuclear systems
 - 10 The theory of diatomic molecules
 - 11 Scattering theory

- 12 The method of second-quantization and radiation theory
- 13 Relativistic quantum mechanics
- 14 Some problems of quantum electrodynamics
- 15 Fundamentals of the theory of elementary particles

Subject index

Volume 4

Part VI Quantum Statistics and Physical Kinetics

- Chapter 1 Quantum statistics
- 2 Physical kinetics
- 3 Kinetic theory of gases and gas-like systems
- 4 Time correlation function method and Onsager's theory
- 5 Solid-state theory

Subject index

Contents of Volume 2

Part III Statistical physics

Chapter 1 The basic concepts of the theory of probability

§ 1 Problems of statistical physics. Necessary results from classical and quantum mechanics	1
2 Basic ideas of the theory of probability	19
3 Mean values and fluctuations	25
4 Normal distribution and moments	31
5 The correlation function	35

Chapter 2 The kinetic theory of gases

6 The simplest statistical system — an ideal gas	39
7 The Maxwell distribution	43
8 Collisions of molecules with the wall of the container. Pressure. The connection of the parameter α with the absolute temperature	47
9 Properties of the Maxwell distribution	51
10 The calculation of characteristic quantities	55
11 Collisions of molecules with each other	59
12 The mean free path	62

Chapter 3 Statistical distribution

13 Quasi-independent systems	66
14 Statistical distribution	68
15 The probability of a state of a system	70
16 The Gibbs distribution	75
17 The statistical temperature	82
18 The properties of the Gibbs distribution and statistical equilibrium	84

§19	Transition to classical statistics	86
20	The monatomic gas as a whole	91

Chapter 4 Statistical and phenomenological thermodynamics

21	The internal energy of a macroscopic system. The first and second laws of thermodynamics	97
22	Work and pressure	101
23	The change in the energy of a system in the general case of a quasi-static process	103
24	Entropy and the basic thermodynamic equality	108
25	The law of increase of entropy	111
26	The basic thermodynamic inequality	116
27	The maximum work to be obtained from thermal processes. The impossibility of constructing a perpetual motion machine of the second kind. Phenomenological definition of the entropy	118
28	The maximum work in non-cyclic processes and the thermodynamic potentials	124
29	Properties of the thermodynamic potentials	127
30	Some thermodynamic relations	130
31	Methods for the transformation of thermodynamic quantities	132
32	The determination of thermodynamic quantities by the methods of statistical physics	137
33	The determination of thermodynamic quantities from experimental data	141
34	Throttling	144
35	The third law of thermodynamics	147
36	The statistical character of the second law of thermodynamics	153

Chapter 5 Ideal gases

37	The distribution function for ideal gases	164
38	The Maxwell-Boltzmann distribution and the Boltzmann distribution in a uniform field of force	174
39	The calculation of the heat capacity of diatomic molecules by means of classical statistics. The law of equipartition of energy over degrees of freedom	179

§40	The thermodynamic functions of a system which can be in two quantum states	190
41	Diatomic molecules	194
42	Thermodynamic functions of diatomic gases	200
43	The vibrational partition function and the contribution of vibrations to the energy and heat capacity	202
44	The rotational partition function and the contribution of rotation to thermodynamic functions	208
45	Polyatomic molecules	213
 <i>Chapter 6 Systems of interacting particles</i>		
46	Interaction between molecules in non-ideal gases	220
47	The correlation function method and its application to the theory of dense gases and liquids	230
48	Equations for correlation functions	232
49	The equation of state and the energy of a system of interacting particles	235
 <i>Chapter 7 Crystals</i>		
50	Crystal structure and thermal motion in the one-dimensional crystal model	242
51	Long waves in a three-dimensional crystal	256
52	The partition function of a crystal	260
53	The thermodynamic functions of a crystal	263
54	Comparison of theory with experiment	265
 <i>Chapter 8 The theory of fluctuations</i>		
55	Small fluctuations in macroscopic systems	270
56	Brownian motion	276
57	Fluctuations of thermodynamic quantities in a homogeneous system	284
58	Effect of fluctuations on the sensitivity of measuring devices	290
 <i>Chapter 9 Systems with a variable number of particles</i>		
59	The Gibbs grand canonical distribution	294
60	The basic thermodynamic equality and the calculation of chemical potentials	301
61	Conditions of phase equilibrium	305

§ 62	The equation of the phase equilibrium curve. Equilibrium between the vapour and the condensed phase	307
63	Theory of phase transitions	315
64	Phase equilibrium curves	320
65	Surface tension and surface pressure	325
66	Gas adsorption	331
67	Chemical equilibrium in the gas phase	334
68	The law of mass action	336
69	Thermal dissociation of atoms	340

Chapter 10 Statistical distributions in quantum statistics and some of their applications

70	The identity of elementary particles and the calculation of the partition function	343
71	A second method of deriving the statistical distribution	345
72	Quantum distributions for an ideal gas	348
73	Black-body radiation	356
74	The classical theory of black-body radiation	360
75	The Planck formula	362
76	The statistics of the photon gas	365
77	The properties of liquid helium II	371
78	Statistical theory of liquid helium II	375
79	The electron gas at absolute zero	381
80	The electron gas at low temperatures	385

Part IV Electromagnetic processes in matter

Chapter 1 Electromagnetic fields in matter

§ 1	The derivation of the basic field equations	397
2	The polarization of a medium in an electric field	400
3	The mean current density and mean charge density in a medium	403
4	The system of equations for the electromagnetic field in a medium	409
5	Boundary conditions	413

§ 6 The limits of applicability of the system of constitutive equations	416
7 The law of conservation of energy in a medium	419

Chapter 2 Electrostatics

8 The electrostatic field	422
9 The solution of electrostatic problems	427
10 The method of images and method of inversion	430
11 The energy of a system of conductors	435
12 Dielectrics and conductors in an external electrostatic field	437
13 The thermodynamic potentials of a dielectric and the dielectric susceptibility	441

Chapter 3 Direct electric current and the magnetic properties of matter

14 Ohm's law	452
15 A linear conductor carrying a constant current	455
16 Direct current in a conducting medium	457
17 The magnetic fields of direct currents. The Biot—Savart law	460
18 The magnetization of magnetic materials and the magnetic moment	464
19 Paramagnetic susceptibility	471
20 Ferromagnetism. Spontaneous magnetization and hysteresis	476
21 Superconductivity	485

Chapter 4 Quasistationary electromagnetic fields

22 Conditions of quasistationarity	493
23 The law of induction in moving conductors and media	496
24 Maxwell's equations for quasistationary fields in integral form and their integration for the case of linear conductors	499
25 The energy of the magnetic field of a system of quasistationary currents	505
26 Coefficients of self-inductance and mutual inductance for non-linear conductors	508
27 Lagrange's equations for a system of quasistationary currents	512
28 The generalized ponderomotive forces in a system with moving current loops	517

§29 Fluctuations in conductors and the Nyquist formula	523
30 Skin effect	528

Chapter 5 High-frequency fields

31 Electromagnetic waves in a homogeneous isotropic medium	533
32 Dispersion relations (the Kramers—Kronig formulae)	540
33 The electromagnetic field in a medium with spatial and time dispersion	544
34 The dispersion of light	549
35 Geometrical optics	555
36 Diffraction	560
37 The reflection and refraction of electromagnetic waves at the boundary between media	570
38 Wave-guides	577
39 The passage of fast particles through matter	585

Chapter 6 Matter in the plasma state

40 The general characteristics of plasma	595
41 Equilibrium plasma	597
42 Plasma in static electric and magnetic fields	605
43 Magnetic isolation and the pinch effect	610
44 The magnetic field in a moving plasma	613
45 Magnetohydrodynamic waves	617
46 Plasma in a high-frequency electric field	622

Appendix IV Important integrals. Stirling's formula	629
---	-----

Subject index	633
---------------	-----

PART III

STATISTICAL PHYSICS

The Basic Concepts of the Theory of Probability

§1. Problems of statistical physics. Necessary results from classical and quantum mechanics

In this part we shall acquaint ourselves with the fundamentals of the atomic theory of macroscopic bodies. By macroscopic bodies we shall understand systems made up of a very large number of particles. It is usual to divide the atomic theory of macroscopic bodies into two parts — statistical physics and physical kinetics.

In statistical physics one restricts oneself to the treatment of the properties of macroscopic systems whose states do not change in time. The states in which a macroscopic system can remain for an indefinitely long time are called equilibrium states. Hence it can be said that the problem of statistical physics (sometimes called statistical mechanics or physical statistics) is the investigation of the properties and behaviour of macroscopic systems in a state of equilibrium on the basis of the known properties of the particles constituting them.

Macroscopic bodies can be made up of elementary particles — electrons, protons, neutrons and so on, or their combinations — nuclei, atoms and molecules. For brevity we shall call all the particles, i.e. molecules, atoms, electrons, protons etc., microparticles. The major constituents of bodies in

ordinary physical conditions are the atoms or molecules representing their structural units. Only in a high-temperature plasma (see Part IV) does one need to take into account the possibility of the dissociation of atoms into electrons and nuclei.

In statistical physics the properties and the laws of motion of the elementary particles, atoms and molecules, are considered known. The problem consists of finding the behaviour of systems containing very large numbers of particles with known properties.

The investigation of the properties of such macroscopic systems by the methods of statistical physics allows one to show an essential feature of such systems. This is that the behaviour of macroscopic systems is determined by laws of a special type, statistical laws.

It turns out that the general equilibrium properties of systems depend relatively little on the actual properties of the particles constituting the bodies and their laws of interaction. Hence in statistical physics one succeeds in establishing the general laws of behaviour of all macroscopic bodies in a state of equilibrium. In particular, statistical physics allows one to find the universal laws of thermal behaviour of macroscopic bodies (the laws of thermodynamics). However, the application of a number of general relations of statistical physics to concrete systems requires a certain amount of information, albeit very restricted, about the laws determining the behaviour of atomic systems.

We have stressed more than once that classical physics turns out to be inapplicable in the realm of atomic phenomena. Hence the application of the laws of statistical physics to real systems is, in practice, impossible if one tries to restrict oneself to the classical concepts of the motion of atomic particles. Therefore we shall be obliged to present a certain amount of data from quantum mechanics.

In addition to establishing the equilibrium properties of macroscopic bodies it is also of great interest in physics to find the behaviour of bodies whose states change in time. The study of the non-equilibrium properties of macroscopic systems is the problem of physical kinetics. It is clear that the laws of change of state of macroscopic systems — the laws of physical kinetics — are substantially more complex than the laws determining the behaviour of equilibrium systems. In physical kinetics it has not been possible to obtain rigorous laws for the change of state of systems in time which have universal validity. Hence at present there exist only laws for the behaviour of non-equilibrium systems of the simplest character.

We now turn to the presentation of the results required from classical and quantum mechanics. For a clear description of the behaviour of mechanical

systems, the motion of which is described by Hamilton's equations

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad \dot{q}_i = \frac{\partial H}{\partial p_i} \quad (i=1, 2, \dots, f) \quad (1.1)$$

(where p_i and q_i are respectively the generalized momenta and generalized coordinates, f is the number of degrees of freedom equal to $3N$, N is the number of material points in the system, and H is the Hamiltonian), use is often made of graphical methods. One of these methods is the motion of a state of a mechanical system in phase space.

This is the space in which the generalized coordinates and momenta are chosen as the coordinate axes.

Consider first the case of a system with one degree of freedom. Let the dependence of the coordinate q and momentum p on time be known. Then graphs of $q(t)$ and $p(t)$, showing the change of these quantities in time, can be plotted. It is more convenient, however, to have a graph representing a sequence of the states of a system, than to have individual graphs presenting the change of the position and momentum of the system. In order to obtain the graph of a sequence of states it is necessary to match the two graphs $q(t)$ and $p(t)$, eliminating the time from them. We choose the generalized coordinate q as the x -axis, and the generalized momentum p as the y -axis. The curve in fig. III.1 shows the change of states of the system. Thus, for example, at point 1 the system had a coordinate q_1 and a momentum p_1 , at point 2 it had a coordinate q_2 and a momentum p_2 and so on. As the time increases the coordinate q of the system and its momentum vary according to the law shown by the graph. The space of fig. III.1 is called phase space. It should be strongly emphasized that phase space has nothing in common with real space and is a purely imaginary concept.

To each point of phase space there corresponds a definite state of the system. A point whose position in phase space characterizes a state of the system is called the representative point. As the state of the system changes,

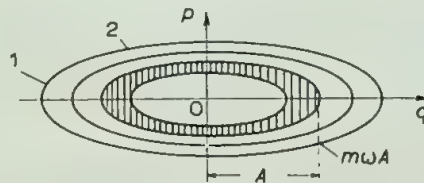


Fig. III.1

i.e. as the position and momentum of the system vary in real space, the position of the representative point in phase space changes and the point describes a certain phase trajectory. The form of this trajectory is not at all like that of the real trajectory. However, they are connected, as well as with the law of change of the momentum, by a one-to-one correspondence.

In order to picture the above more clearly, we shall consider the motion of a linear harmonic oscillator moving under the action of a quasi-elastic force $F = -\kappa q$ about the origin $q = 0$. The equation of motion has the form

$$m\ddot{q} = -\kappa q. \quad (1.2)$$

It is easily integrated. We have

$$q = A \sin(\omega t + \alpha). \quad (1.3)$$

In this case the momentum of the oscillator is equal to

$$p = m\omega A \cos(\omega t + \alpha). \quad (1.4)$$

where A is the amplitude and α is the phase, determined by the initial conditions; the frequency is

$$\omega = 2\pi\nu = \sqrt{\kappa/m}.$$

Formula (1.3) represents the equation of the real motion. In order to find the trajectory of the representative point in phase space it is necessary to find the relation between p and q . Squaring and adding up eqs. (1.3) and (1.4), we find

$$\left(\frac{p}{m\omega A}\right)^2 + \left(\frac{q}{A}\right)^2 = 1.$$

This is the equation of an ellipse. Thus, for oscillations about the point $q = 0$ with an amplitude A the representative point in phase space describes an ellipse with semi-axes $a = A$ and $b = m\omega A$ (fig. III.1). Its area S is equal to $S = \oint p dq = \pi ab = \pi m\omega A^2$. Also, calculating the energy of the oscillator ϵ , we have

$$\epsilon = \frac{p^2}{2m} + \frac{1}{2}\kappa q^2 = \frac{1}{2}\kappa A^2, \quad (1.5)$$

from which we find the important relation

$$\epsilon = \frac{S_K}{2\pi m\omega} = \nu \oint p \, dq . \quad (1.5')$$

The notion of phase space can also be introduced for a system with more than one degree of freedom. In this case the number of dimensions of the phase space is obviously equal to twice the number of degrees of freedom, since for each degree of freedom the coordinate is taken as one axis, and the momentum is taken as another axis. In the case of systems with a large number of degrees of freedom the phase space has very many dimensions and cannot be shown graphically. Nevertheless, in this case the concept of phase space also proves to be of great use. In what follows we shall need the expression for a volume element in phase space. Generalizing the usual definition of a volume element $dV = dx dy dz$ to the case of many dimensions, one can write the following expression for an element of phase space

$$d\Gamma = dq_1 dq_2 \dots dq_{3N} dp_1 dp_2 \dots dp_{3N} , \quad (1.6)$$

where dq_i is the differential of the i th coordinate, and dp_i is the differential of the i th momentum corresponding to this coordinate ($i = 1, 2, \dots, 3N$). The product on the right-hand side of the expression (1.6) involves the differentials of $3N$ generalized coordinates and $3N$ momenta as factors.

In addition to the idea of phase space, use is often made of the ideas of configuration space and momentum space.

The representative space of $3N$ dimensions, in which generalized coordinates are chosen as the axes, is called the configuration space. The set of all positions of the particles of a system is characterized by the position of a representative point in the configuration space. An element of the configuration space is

$$dV_{\text{conf}} = dq_1 \dots dq_{3N} . \quad (1.7)$$

In momentum space the $3N$ components p_1, p_2, \dots, p_{3N} of the momenta of the particles serve as the coordinate axes. A representative point characterizes the value of all the momenta of the particles. An element of the momentum space is equal to

$$dV_{\text{mom}} = dp_1 \dots dp_{3N} , \quad (1.8)$$

so that

$$d\Gamma = dV_{\text{conf}} dV_{\text{mom}}. \quad (1.9)$$

The notion of phase space, as well as those of configuration and momentum space, are very useful for a clear representation of the laws of statistical physics.

The motions of mechanical systems are determined by the so-called dynamical laws. A characteristic feature of a dynamical law is that, if the initial state of the system and the action of surrounding bodies on the system are known, the state of the system at any subsequent instant of time can be determined unambiguously. In other words, for given forces acting on the system the initial state of the system unambiguously determines its subsequent overall motion.

General features characteristic of a dynamical law show up not only in mechanics but also in a wide range of other physical phenomena, in particular in electrodynamics. However, it would in principle be incorrect to state, as was done by a number of investigators, beginning with Laplace, that dynamical laws exhaust all forms of causality and mutual conditionality of phenomena in nature.

As we shall see below, the behaviour of macroscopic bodies is not given by dynamical laws but is determined by laws of another type — statistical laws.

In the following we shall assume that atoms, molecules and other particles forming macroscopic systems move according to the laws of quantum mechanics. The latter are presented in Part V of this book. Here we shall give, without proof, the most necessary data and relations.

A distinctive feature of all microsystems (atoms, molecules, etc.) is the fact that in certain conditions they can be in discrete, or quantum, states. Experimental and theoretical (see Part V) investigation of the states of atoms and molecules has shown that their energy can assume a discrete series of values $\epsilon_1, \epsilon_2, \epsilon_3, \dots$. The transition between these states, for example, between ϵ_1 and ϵ_2 , takes place without passing through states with intermediate energies between ϵ_1 and ϵ_2 . States with intermediate energies do not exist. Thus, atoms can absorb or emit energy only in definite amounts or quanta.

Not only the energy but also a number of other quantities characterizing the states of atomic systems have a discrete, quantum, character (for example, the angular momentum, which in the atom can assume a discrete series of values and can vary only in jumps). The energy and similar quantities are called quantized quantities, and the set of their possible values is called the spectrum. The quantized values of energy are also often called energy levels.

The existence of quantized states radically contradicts the laws of classical mechanics, in which the states of a system can always change continuously.

At the beginning of the development of atomic theory some formal rules were obtained, by means of which it was possible to choose from all states allowed according to classical mechanics, those which can actually be realized in an atom. These rules were called Bohr's quantum conditions.

Subsequent development of atomic physics showed that the concepts of classical physics needed an even more profound alteration.

At present the laws of motion of microparticles are adequately described by quantum mechanics. Postponing our detailed acquaintance with them till Part V, we shall, in the meantime, make use of some of them.

1.1. *Quantized motion of a particle in a box*

The simplest quantum-mechanical system – a microparticle with mass m confined in a one-dimensional box with impenetrable walls – will be considered in §8 of Part V. The motion of such a particle is restricted by the size of the box to the region $0 \leq x \leq a$, where the potential energy is equal to zero. At the limits of the region, at $x = 0$ and $x = a$, the potential energy of repulsion is infinitely large and the particle cannot go out of the box.

From the general propositions of quantum mechanics it follows that the energy and momentum of such a particle should have a discrete series of values. The allowed values of the energy and momentum are given by the formulae

$$p_n = \frac{nh}{2a}, \quad (1.10)$$

$$\epsilon_n = \frac{p_n^2}{2m} = \frac{h^2 n^2}{8ma^2}. \quad (1.11)$$

Here h is a quantity called the universal quantum constant or Planck constant. It is equal to $h = 6.62 \times 10^{-27}$ erg sec. n is a quantity running through the integer values ($n=1, 2, 3, \dots$), and called the quantum number.

Formula (1.11) shows that the energy levels of the particle form a discrete series or spectrum. The separation between neighbouring energy levels is equal to

$$\Delta\epsilon_n = \epsilon_{n+1} - \epsilon_n = \frac{h^2}{8ma^2} (2n+1). \quad (1.12)$$

We see that these separations are smaller the larger the mass of the particle and the dimension a of the region of motion.

In the case of the motion of a particle in a region of sufficiently large dimensions the separation between energy levels is so small that for practical purposes they form a continuous spectrum. Similarly any particle with a large, macroscopic mass has a continuous energy spectrum.

The relative separation between energy levels is obviously equal to

$$\frac{\epsilon_{n+1} - \epsilon_n}{\epsilon_n} = \frac{2n+1}{n^2}. \quad (1.13)$$

For $n \gg 1$ the relative separation between energy levels or the magnitude of the "steps" of the energy spectrum is equal to

$$\frac{\epsilon_{n+1} - \epsilon_n}{\epsilon_n} \approx \frac{2}{n}. \quad (1.14)$$

For large quantum numbers the relative separation between energy levels decreases rapidly with increasing n , so that the discrete character of the spectrum is smoothed away.

Thus we see that the discreteness of the energy levels of a quantum particle becomes apparent: 1) for small mass, 2) for motion of the particle in a small region, and 3) for small quantum numbers. On the contrary, for large masses, for motion in a large region, and for large quantum numbers the quantization shows up relatively weakly.

In order to have a feeling for the orders of magnitude, let us put in some numbers. For example, let a proton with mass $m_p = 1.7 \times 10^{-24}$ g move in a box which has atomic dimensions ($a = 10^{-8}$ cm). Expressing the energy in electron-volts we find

$$\epsilon_n = 0.02n^2 \text{ eV}$$

and

$$\Delta\epsilon_n = 0.02(2n+1) \text{ eV}.$$

For small n 's the separations between energy levels turn out to be of the same order of magnitude as the energies themselves (for example, for $n = 3$, $\epsilon_3 \approx 0.2$ eV, $\Delta\epsilon_3 \approx 0.1$ eV).

However, the situation is different when a proton moves in a region of macroscopic dimensions (for example, $a = 1$ cm). Then

$$\epsilon_n = 2 \times 10^{-18} n^2 \text{ eV} \quad (1.15)$$

and the separation between neighbouring levels is

$$\Delta\epsilon_n = 2 \times 10^{-18} (2n+1) \text{ eV} . \quad (1.16)$$

Let the proton have an energy of 2×10^{-2} eV (as will be seen in what follows such an energy is possessed by atoms in thermal motion at normal temperature). Then from (1.15) we find $n \approx 10^8$. For such values of n the relative separation between energy levels is negligibly small. Thus, when the proton moves in a region of sufficiently large dimensions the discrete, quantum character of its states is manifested very weakly. The same holds, to an even higher degree, for a macroscopic ball with a mass of, say, 1 g. The motion of such a ball is described with a high degree of accuracy by the laws of classical mechanics.

The regularities appearing in the special case which we have considered, of a particle moving in a box, are of general character.

In Part V it will be shown that classical mechanics represents the limiting case of quantum mechanics into which the latter goes over when effects proportional to the Planck constant can be disregarded. This is possible for phenomena which take place on a relatively large scale, when the masses of the particles, the dimensions of the region of motion and so on are sufficiently large.

It turns out that the transition from quantum mechanics to classical mechanics can be made in two ways. Simply, by assuming $\hbar = 0$, we disregard all quantum effects — the existence of the wave properties of microparticles, the quantization of energy and other quantities, and so on. However, it can also be assumed that \hbar is small but nevertheless different from zero. It turns out that in this approximation the wave properties of particles are shown very weakly. Particles can be assumed to be moving on definite trajectories, as in classical mechanics. However, the quantization of states still arises in the fact that only some of the classical trajectories are possible. Such an approximation is called quasi-classical (as distinct from the classical approximation, in which the quantum properties of particles are not taken into account at all). In order to see what the character of the restrictions imposed upon the classical trajectories consists of, we again turn to the example of a particle in a one-dimensional box. We assume that use can be made of the concepts of classical

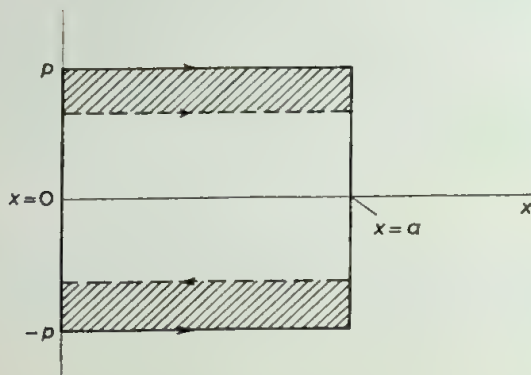


Fig. III.2

mechanics and that the particle can be considered as a material point moving between reflecting walls. The phase diagram in fig. III.2 shows the sequence of its states. We now take into account the quantization of states and out of all possible states we choose those which satisfy the quantization condition (1.10). Not all states $p = \text{const}$ are possible, only those with a separation determined by the relation (1.10). Fig. III.2 shows the n th state (the solid line) and the $(n-1)$ th state (the dotted line). The number of possible quantum states with a momentum lying between $p_n = hn/2a$ and $p = h/2a$ is equal to n .

We now calculate the area S_n in the phase plane, corresponding to these n states. Obviously,

$$S_n = \oint p \, dx = 2p_n a = hn.$$

The integral $\oint p \, dx$ denotes the integral of p taken over the total cycle of motion, i.e. over the area bounded by the solid straight lines in fig. III.2. This integral is equal to

$$\oint p \, dx = \int_0^a p \, dx - \int_a^0 p \, dx = 2 \int_0^a p \, dx.$$

If the lines corresponding to other possible states are drawn, then the entire surface will be divided into cells. It is easily seen that the area of all the cells is the same and equal to h . Indeed, the distance between possible states along the p -axis is equal to

$$\frac{hn}{2a} - \frac{h(n-1)}{2a} = \frac{h}{2a}.$$

The area of a cell (hatched in the drawing) is equal to $2(h/2a)a = h$.

Thus, in the quasi-classical approximation to each possible state there corresponds in phase space a cell having an area h . The stationary, possible states of a system are those for which the condition

$$\oint p \, dx = nh \quad (1.17)$$

is fulfilled. This equality is called the Bohr condition of the old quantum theory.

The example considered is typical, and the condition (1.17) is of a general character. In order to convince ourselves of this we shall consider another example — a linear oscillator. A diatomic molecule is an example of an oscillator performing small vibrations about the equilibrium position, as we shall see below. In quantum mechanics (see § 10 of Part V) it turns out that the state of an oscillator is characterized by a quantum number k which can take on a number of half-integer values

$$k = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots, \quad k = n + \frac{1}{2}$$

(n is an integer). The energy of an oscillator takes on a number of values (see (10.13) of Part V)

$$\epsilon_n = h\nu \left(n + \frac{1}{2}\right). \quad (1.18)$$

When an oscillator passes from a given quantum state to a neighbouring, lower, one it emits radiation with a frequency

$$\nu_{mn} = \frac{\epsilon_m - \epsilon_n}{h} = \nu,$$

equal to the natural frequency of oscillation of the classical oscillator*.

Comparing (1.18) with formula (1.5') we see that the quantum condition (1.18) selects as possible those states of an oscillator for which the relation

* In quantum mechanics it turns out that for an oscillator only the transitions between neighbouring states are possible, so that $m = n + 1$. See Part V.

$$\oint p \, dq = h(n + \frac{1}{2}) . \quad (1.19)$$

holds.

All possible paths of the representative point of the oscillator, corresponding to quantum states n_1, n_2, \dots , are represented by ellipses. The area of the ellipse corresponding to a state n differs from that of the ellipse corresponding to a state $n-1$ by an amount

$$\oint_n p \, dx - \oint_{n-1} p \, dx = h ,$$

where the subscript denotes the number of the state. In fig. III.2 this area is hatched.

We arrive at the conclusion that a cell in phase space the area of which is equal to h corresponds to each quantum state of an oscillator.

Thus, in the quasi-classical approximation (for large quantum numbers or dimensions of the region of motion, and for large particle masses) the condition of quantization of states consists of the fact that to each quantum state of an arbitrary system there corresponds a cell in phase space having an area h . It can be shown that the form of the cell is arbitrary.

Up to now we have confined ourselves to the consideration of systems having one degree of freedom. However, it turns out that the results obtained are of general character and can be applied to a system with an arbitrary number f of degrees of freedom. The state of such a system is characterized by the definition of f quantum numbers (for examples see below). The phase space of a system with f degrees of freedom has $2f$ dimensions.

Again as an illustrative example we shall consider the motion of a free particle in a box with ideally reflecting walls, but now having three dimensions. For simplicity we assume the box to have the form of a cube with edge a . Since the motion in any of the three directions is independent and all the directions are in principle equivalent, we can write for each of the components of the momentum

$$|p_x| = \frac{hn_1}{2a} , \quad |p_y| = \frac{hn_2}{2a} , \quad |p_z| = \frac{hn_3}{2a} . \quad (1.20)$$

The motion of a particle in the three dimensions is characterized by three quantum numbers n_1, n_2, n_3 which can take on a number of integer values. The energy of the particle is equal to

$$\epsilon = \frac{p_x^2 + p_y^2 + p_z^2}{2m} = \frac{h^2}{8ma^2} (n_1^2 + n_2^2 + n_3^2). \quad (1.21)$$

It is characterized by a number $n = \sqrt{n_1^2 + n_2^2 + n_3^2}$, but for given n does not depend on the value of the separate contributions to this n of each of the quantum numbers n_1, n_2, n_3 . Because of this, several different quantum states can correspond to one and the same value of the energy. For example let $n_1 = 1, n_2 = 2, n_3 = 2$ and $n_1 = 2, n_2 = 1, n_3 = 2$. In both cases $n = 3$, so that both states have the same energy. If one and the same energy corresponds to several different states, then such states are called degenerate.

The number of states having the same energy is called the degeneracy or statistical weight.

The phase space of a particle in a box has six dimensions, so that it is impossible to present it graphically. However, it can be said that it is split into three subspaces of two dimensions, each corresponding to the motion in the corresponding direction. A simple calculation leads us then to the conclusion that to each state (the triplet of numbers n_1, n_2, n_3) of the particle there corresponds a volume h^3 .

In the most general case of an arbitrary system having f degrees of freedom it can be shown that in the transition to the quasi-classical approximation, the motion of the system can be considered in the same way as in classical mechanics, but with a restriction imposed on the possible states: to each quantum state of a system with f degrees of freedom in the quasi-classical approximation there corresponds a cell in its phase space having a volume h^f . The proof of this statement will be given in §41 of Part V.

In the exposition of statistical physics we shall need, in most cases, to consider the motion of relatively heavy particles (for example, molecules) moving in macroscopic volumes, as well as the behaviour of macroscopic bodies containing an enormous number of molecules.

For such systems quantum effects play a relatively minor role. Nevertheless, as will be seen later, they cannot be completely disregarded. Hence we shall take them into account in the quasi-classical approximation, based on the quantization rule quoted above. In other respects, if not specified otherwise, the motion of the systems will be considered classically. Of course, in certain cases when the mass of the system is sufficiently large, one can pass over from the quasi-classical treatment to the purely classical one and completely disregard quantum effects. This will be done in Chapter 3. However, in general discussions and conclusions we shall assume the states of the system to be discrete.

1.2. Number of quantum states

The notion of the number of quantum states corresponding to the energy of a system lying in a given interval between ϵ and $\epsilon + \Delta\epsilon$, will play an important role in our overall further development. We shall denote it by $\Omega(\epsilon)\Delta\epsilon$.

We shall first calculate this number for a particle moving freely in a box. According to the above, there corresponds to each state a volume h^3 of the phase space. Hence the number of states sought for will be found by calculating the volume of the phase space corresponding to the energy of the particle lying between ϵ and $\epsilon + \Delta\epsilon$, and dividing it by h^3 .

In calculating the phase volume we shall make use of the fact that in the quasi-classical approximation quantum jumps are small, and we shall assume the momentum to vary almost continuously. Then an element of the phase space can be written in the form (1.6). The volume of the phase space corresponding to an energy of the particle smaller than a given value is obtained by integrating the expression (1.6) with respect to all coordinates and all momenta satisfying the relation $0 \leq p \leq \sqrt{2m\epsilon}$. Passing over to spherical coordinates, we can write

$$\begin{aligned} \Gamma &= \int dx dy dz \int dp_x dp_y dp_z \\ &= 4\pi V \int_0^{\sqrt{2m\epsilon}} p^2 dp = \frac{4\pi p^3 V}{3} \Big|_0^{\sqrt{2m\epsilon}} = \frac{4\pi(2m)^{\frac{3}{2}} \epsilon^{\frac{3}{2}} V}{3}. \end{aligned} \quad (1.22)$$

The volume of phase space corresponding to energies in the range between ϵ and $\epsilon + \Delta\epsilon$ is

$$\Delta\Gamma = \frac{\partial\Gamma}{\partial\epsilon} \Delta\epsilon = 4\pi m V \sqrt{2m\epsilon} \Delta\epsilon. \quad (1.23)$$

The number of states of the particle with energies which lie between ϵ and $\epsilon + \Delta\epsilon$ is equal to

$$d\Omega = \Omega(\epsilon) \Delta\epsilon = \frac{1}{h^3} \frac{\partial\Gamma}{\partial\epsilon} \Delta\epsilon = \frac{4\pi m V \sqrt{2m\epsilon}}{h^3} \Delta\epsilon. \quad (1.24)$$

Since all quantities in the quasi-classical approximation vary almost continuously, we shall often write $\delta\epsilon$ instead of $\Delta\epsilon$ in formula (1.24), assuming $\delta\epsilon$ to be infinitely small.

It should be borne in mind that for large values (large quantum numbers)

the number of states corresponding to even a very small interval $\Delta\epsilon$ turns out to be enormous. Thus, for example, in the case of hydrogen atoms for $\Delta\epsilon = 0.005$ eV, $\epsilon = 0.025$ eV and $V = 1$ cm³ the value of $\Omega \Delta\epsilon$ turns out to be about 4×10^{28} . This value in practice differs very little from its classical limit i.e. infinite (as $\hbar \rightarrow 0$). Nevertheless, the finiteness of the number of quantum states, as we shall see later, plays an important role.

For an arbitrary system having f degrees of freedom

$$\Delta\Gamma = \frac{\partial\Gamma}{\partial\epsilon} \Delta\epsilon, \quad (1.25)$$

where $\Delta\Gamma$ is determined by formula (1.6). Correspondingly, for the number of states we have

$$\delta\Omega = \Omega(\epsilon)\delta\epsilon = \frac{1}{h^f} \frac{\partial\Gamma}{\partial\epsilon} \delta\epsilon, \quad (1.26)$$

or

$$\Omega = \int d\Omega = \Delta\Gamma/h^f. \quad (1.26')$$

The quantity $\Omega(\epsilon)$ can be called the number density of states per unit energy range. In what follows we shall, for brevity, simply call $\Omega(\epsilon)$ the number of states with given energy. This should not lead to misunderstanding.

For what follows we shall need one more sufficiently obvious property of Ω . That is, if there is a system consisting of two independent parts, and the number of states of each of the parts is equal to Ω_1 and Ω_2 respectively, then the number of states of the combined system is equal to $\Omega = \Omega_1 \Omega_2$. Indeed, the phase volume of a combined system is by definition equal to $d\Gamma = d\Gamma_1 d\Gamma_2$, from which the above mentioned property follows immediately.

In the general case

$$\Omega = \prod_i \Omega_i, \quad (1.27)$$

where the product \prod_i is taken with respect to all parts of the system.

We shall make use of this property of Ω in order to estimate the number of states of a system consisting, for example, of 100 independent particles moving in a volume $V = 1$ cm³ with an energy within the interval $\Delta\epsilon = 0.005$

eV, for $\epsilon = 0.025$ eV and a mass equal to that of the proton. We find

$$\Omega = \Omega_1 \Omega_2 \dots = (4 \times 10^{28})^{100} \approx 10^{2860}.$$

1.3. Spin

Up to now, in considering an individual microscopic particle (for example, an electron or proton), we have assumed that its state is completely characterized by the definition of three quantum numbers corresponding to three degrees of freedom. However, it turns out that for a complete characterization of the state of an elementary particle one more quantum number must be specified.

A large number of particles possess, besides an orbital angular momentum, an additional, intrinsic angular momentum which is not associated with a spatial displacement. It is called the spin. The smallest angular momentum which can be possessed by an elementary particle is its spin. Most elementary particles (electrons, neutrons, protons) possess a spin s equal to $\frac{1}{2}\hbar/2\pi$. This means that the spin projection s_z onto an arbitrary z -axis in space can have two values: $\frac{1}{2}\hbar/2\pi$ and $-\frac{1}{2}\hbar/2\pi$. It is said that the spin coordinate takes on two values: $\frac{1}{2}$ and $-\frac{1}{2}$. The spin of complex particles can be integer as well as half-integer, depending on the elementary particles constituting them.

1.4. The principle of identity of elementary particles

It turns out that the idea of the discrete character of the states of a system and, in particular, discrete energy levels, allows one to treat a wide range of problems which were unsolved in classical physics. However, in addition it will be necessary also to take into account some other features of quantum systems which basically affect the behaviour of real macroscopic systems. In the following we shall often have to deal with systems consisting of numbers of identical particles (for example, electrons or atoms of given type). The laws of the behaviour of such systems in quantum mechanics differ sharply from classical laws. In classical physics, however similar physical bodies may be in their properties, in principle one can always follow their individual motion and distinguish them from one another.

In quantum mechanics the situation is radically different. The reason for this lies in the fact that in quantum mechanics the principle of identity of like particles holds. According to this principle all particles of a given kind (for example electrons) entering into a given quantum-mechanical system are completely identical. In a system consisting of particles of a given kind the states do not change when particles are interchanged.

For example, let the system consist of two electrons. The first electron is in a state characterized by a set of quantum numbers n_1 , and the second electron is in a state with quantum numbers n_2 . If the states of these electrons are exchanged, then we obtain a state of the system with the same energy. At first sight the impression that the states of the system are two-fold degenerate may arise. However, from arguments based on the general propositions of quantum mechanics as well as from statistical considerations it is possible to state that this is not so (see §37 and §64 of Part V).

The identity of particles of a given kind is so complete that, for example, the exchange of an electron in a given state by another is not a physical event. Hence it makes no sense to say that electron No. 1 is in a state 1 and electron No. 2 is in a state 2. It should be noted that a system of two electrons is in a definite state. From this statement, which follows directly from a number of experimental facts, very important consequences for statistical physics are obtained. We shall acquaint ourselves with these in Chapter 5, and, in particular, in Chapter 10.

The properties of systems of particles with integer spin differ so fundamentally from those of systems of particles with half-integer spin that, reasoning strictly, it is necessary to speak of two different aspects of quantum mechanics: one for particles with integer spin and the other for those with half-integer spin. This is apparent from the following. For particles with half-integer spin the so-called Pauli exclusion principle holds: "only one particle with half-integer spin can be in each quantum state".

The exclusion principle is often formulated in a somewhat different way: "no more than two electrons with an opposite spin orientation can be in each quantum state". The equivalence of the two formulations is obvious.

For particles with integer spin there is no restriction upon the number of identical particles occupying a given quantum state. Below it will be shown that this fact radically affects the statistical behaviour of systems consisting of particles with integer spin in comparison with those of half-integer spin.

1.5. *Energy levels of a system consisting of a large number of particles*

Consider a system consisting of N identical atoms or molecules, assuming N to be a large number. From general considerations it is clear that, since the system is macroscopic, the internal energy of the system must vary continuously and quantum effects should not be of essential importance. We shall see now how a continuous energy level distribution arises when atoms with discrete energy levels are combined.

For simplicity we shall consider two hydrogen atoms each in the same non-degenerate state with energy ϵ_0 a large distance apart (in comparison with

their dimensions). At an infinitely large distance the atoms do not interact with each other and the energy of the entire system ϵ is equal to the sum of the energies of the two atoms, i.e.

$$\epsilon = 2\epsilon_0.$$

The state of the system will be two-fold degenerate: the state of the system when one electron is near the first nucleus and the second electron is near the second nucleus will possess the same energy as when the electrons are exchanged, in view of the identity of electrons.

Now draw the atoms nearer to a distance such that they may interact with each other. Calculation shows that when the interaction arises the energy level of the system splits into two levels lying close to each other. It is said that the interaction has removed the degeneracy. This will be treated in detail in §54 of Part V.

If one continues to bring the atoms nearer forming a molecule, the splitting of the levels will increase (fig. III.3, the lower level). Very often the energy levels of each of the atoms are in themselves degenerate. Then there arise from one energy level not two but a larger number of levels of the system of interacting particles (fig. III.3, two upper levels). We see that the number of energy levels in a system of interacting particles turns out to be larger than in a system of separated particles. The degeneracy of the levels is removed by the interaction. This result is not unique to a system of two atoms, but is of general character. If the system is one of atoms, characterized by quantum numbers, then in forming a system of strongly interacting particles, for example a crystal, all the energy levels of individual atoms split into individual energy levels of the system as a whole. The latter, in general, are not degenerate.

If the number of atoms in the system (or, more precisely, the number f) is large, then the total number of energy levels in the system turns out to be enormous. They approach each other rapidly with increasing energy (as is seen in fig. III.3; compare the first, second and third level), and for large f and large excitation energies they merge almost completely, forming continuous bands of allowed energy levels.

From this it is clear that the statement about the continuous variation of the energy of a macroscopic body is not completely true. The lowest energy levels are discrete. As the energy increases energy levels approach each other rapidly and the energy of the system becomes continuous. We shall see later that the discreteness of the lowest energy levels in macroscopic systems fundamentally affects their behaviour at very low temperature, close to absolute zero.

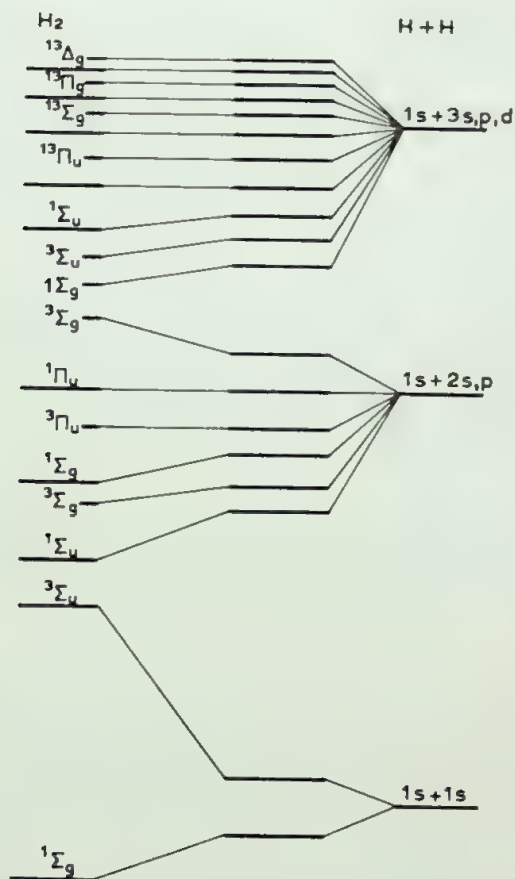


Fig. III.3

§2. Basic ideas of the theory of probability

Our next problem is the study of statistical laws in systems consisting of a very large number of particles.

This study will be based on the mathematical apparatus of the theory of probability.

We shall not discuss the theory of probability in the form in which it is done in mathematical courses. We shall introduce from the very beginning a

special definition of probability, which is completely equivalent to that adopted in the mathematical theory of probability but is more obvious and convenient in considering probability processes in statistical physics. This definition is closely associated with the concept of the relation between the probability and the frequency of occurrence of an event, used in every-day practice.

Consider a completely arbitrary physical system which can be in different physical states. We assume at first that these states form a discrete series, and number them by the integers 1, 2, 3, We denote by L any quantity depending on the state of the system. The quantity L can represent, for example, the energy, volume, compressibility or any other quantity which is a function of the state and changes with a change of the state of the system. We consider L to be a single-valued function of the state of the system, so that to each state 1, 2, 3, ... there corresponds a well-defined value of the quantity L : L_1, L_2, L_3, \dots . Conversely, if the quantity L has a value L_i , then this means that the system must be in the i th state.

Assume that in the course of a very long time T by virtue of diverse processes taking place in the system its states change in such a way that it passes through a sequence of different states 1, 2, 3, ..., i , For clarity we assume that in the course of the entire time T the value of the quantity L is measured regularly every Δt seconds.

The system will exist in certain states for a long time and will be in them often, whereas other states will be occupied by it for only a negligible time. As a result of the measurements we shall obtain some value of L more often than others. Let the system spend a time t_i , constituting a part of the total observation time T , in a certain state i . As a result of $N_i = t_i/\Delta t$ measurements it will be found that the quantity L has the value L_i . The total number of measurements will obviously be equal to $N = T/\Delta t$. The limit of the ratio of the number of measurements giving a value of L equal to L_i to the total number of measurements, as the latter increases indefinitely, i.e.

$$w_i = \lim_{N \rightarrow \infty} \frac{N_i}{N}, \quad (2.1)$$

will be called the probability of the i th state w_i or the probability of the value L_i . In other words, the probability of the i th state w_i is defined as the limit of the ratio of the time t_i during which the system is in this state to the total observation time T as the latter increases indefinitely:

$$w_i = \lim_{T \rightarrow \infty} \frac{t_i}{T}. \quad (2.2)$$

It is necessary to see clearly that the probability of a given state i and the probability for the quantity L to have the value L_i corresponding to the state i are the same. Hence instead of (2.2) we can write

$$w_{L_i} = \lim_{T \rightarrow \infty} \frac{t_i}{T}. \quad (2.3)$$

where w_{L_i} is the probability for the quantity L to have the value L_i .

The definition (2.2) implies the assumption that the limit of the ratio t_i/T exists. The existence of this limit is ensured in the case where during the course of the entire observation time the system is in unchanged external conditions. If this is not so and the external conditions may vary continuously during the course of the measurements, the ratio t_i/T may not tend to any limit. Thus, for example, if we considered a gas expanding indefinitely, then the system would not be in any state for a finite time interval. Its states would vary continuously during the course of the entire observation time. Hence the limit of the ratio t_i/T would not even exist.

In practice one often encounters systems whose states do not vary in a discrete way but continuously. In other words, the quantities characterizing the state of the system often run through a continuous series of values. In this case the probability definition (2.3) does not have a direct meaning. The system will spend an infinitely short time in a state in which the quantity L has a value exactly equal to L_i . Hence, as in other cases where one has to deal with continuously changing quantities, it is necessary to speak not of the value L_i but of a certain interval of the value of this quantity. We therefore have to speak of the probability for the quantity L to have a value lying in the interval between L and $L + dL$. We shall denote this probability by dw_L . By definition,

$$dw_L = \lim_{T \rightarrow \infty} \frac{\Delta t_L}{T},$$

where Δt_L is the time during the course of which the system is in the states corresponding to the values of L lying between L and $L + dL$. It is obvious that the time Δt_L and, consequently, also the probability dw_L , are proportional to the value of the interval dL , other things being equal. Hence it is convenient to write dw_L in the form

$$dw_L = \rho(L) dL, \quad (2.4)$$

where $\rho(L)$ is the probability for the value of L to lie in a certain "unit" interval. The function $\rho(L)$ is called the probability density. It replaces the probability itself in those cases when the value of L can change continuously.

In addition to the definition (2.2), use is also made in statistical physics of another definition of probability.

Instead of considering the changes of state of a system in time, one can imagine a set of systems which are identical with the given system but which are at a certain instant of time distributed randomly over all possible states. Such a system is called a statistical ensemble. We shall determine the number of systems in an ensemble which are in different possible states. Out of the total number N of systems in the ensemble let N_i systems be in the i th state. Then the probability that a system will be detected in this state in a random measurement is equal to

$$w_i = \lim_{N \rightarrow \infty} \frac{N_i}{N}. \quad (2.5)$$

The probability defined by formula (2.5) is called the probability with respect to the ensemble.

Now let there be a complex mechanical system moving on a certain trajectory in phase space. The complexity of the trajectory rules out the possibility of observing it directly and the phase points are distributed randomly in phase space. The probability of detecting the system in a given region of phase space is, according to (2.1), determined by the time of stay of the system in this region. Instead of observing the process of displacement of the representative point in phase space in time, one can consider an ensemble of systems differing in their initial conditions. If the initial conditions are distributed randomly, then the probability for the system to be found in the i th region of phase space is determined by the number of representative points in this region corresponding to different systems of the ensemble. It is natural to assume that the number of such points for an ensemble is proportional to the time of stay of an individual system in this region. The ratios figuring in the definitions (2.1) and (2.5) lead to the same value of the probability. This assumption in statistical physics is called the ergodic hypothesis. We shall return to a discussion of it in §15.

We shall make use of the two definitions of probability, considering them to be equivalent.

We now consider the definition of certain propositions of the theory of probability.

2.1. *The law of addition of probabilities*

Consider a physical system which can be in different states. If the system is in the state i , then it obviously cannot simultaneously be in any state k . The simultaneous occupation of the states i and k by one and the same system are two events which rule each other out. Assume that the probabilities of the states i and k are known. For many purposes it is very important to find the probability for a system to be found in one or other of these states — no matter in which one. In other words, we want to find the probability for a system to be either in the state i or in the state k . To find this probability we note that the total time of stay of the system in one or other of the states is equal to the sum of the times of stay in the i th and k th state. Hence the probability w_{i+k} sought for is

$$w_{i+k} = \lim_{T \rightarrow \infty} \frac{t_i + t_k}{T} = \lim_{T \rightarrow \infty} \frac{t_i}{T} + \lim_{T \rightarrow \infty} \frac{t_k}{T} = w_i + w_k. \quad (2.6)$$

Formula (2.6) expresses the law (theorem) of addition of probabilities: The probability for a system to be in one or other of two states which are mutually exclusive, is equal to the sum of the probabilities for the system to be in each of the states separately.

The theorem of addition of probabilities can easily be applied to the case of three states or a larger number of states. In the general case the probability for a system to be in one or other of the mutually exclusive states i, k, l, \dots is equal to

$$w_{i+k+l+\dots} = \sum w_j, \quad (2.7)$$

where the summation is carried out over all states i, k, l, \dots of the system.

From the theorem of addition of probabilities there follows an important consequence which we shall often use in what follows.

Assume that the state of a system is characterized by two quantities L and M which are independent of each other. For example, L can represent the velocity of motion of the system in one direction, while M can represent the velocity of motion of the system in another direction, or L and M can be respectively the energy and volume of an ideal gas, and so on. Let L have the possible values $L_1, L_2, \dots, L_i, \dots$, and M the values $M_1, M_2, \dots, M_k, \dots$. We assume that the probability for the system to be in a state in which L is equal to L_i and M is equal to M_k is known. Let this probability be equal to $w_{L_i M_k}$.

We now find the probability w_{L_i} for the system to have the value L_i for any value of the quantity M . According to the theorem of addition of probabilities we can write

$$w_{L_i} = w_{L_i M_1} + w_{L_i M_2} + \dots + w_{L_i M_k} + \dots = \sum w_{L_i M_j}, \quad (2.8)$$

where the summation is carried out over all values of the quantity M . In the case when the quantities L and M vary continuously the summation in formula (2.8) should be replaced by integration.

2.2. Statistical independence and the law of multiplication of probabilities

The second important proposition of the theory of probability is called the theorem or law of multiplication of probabilities.

Consider two physical systems and assume that they are completely independent of each other. We denote by w_{L_i} the probability for the first system to be in the state characterized by the value L_i , and denote by w_{M_k} the probability for the second system to be in the state characterized by the value M_k . The probabilities w_{L_i} and w_{M_k} are independent if the probability for the first system to be in the state i does not depend on whether or not the second system is in the state k .

The law of multiplication of probabilities for statistically independent systems reads: "the probability of the simultaneous occupations of the i th state in which $L = L_i$, by the first system, and the k th state in which $M = M_k$, by the second system, is equal to the product of the probabilities w_{L_i} and w_{M_k} ", i.e.

$$w_{L_i M_k} = w_{L_i} w_{M_k}. \quad (2.9)$$

The law of multiplication represents a strict definition of the statistical independence of two systems.

The above reasoning can be applied to two arbitrary independent physical systems. Let the first of them spend a time Tw_{L_i} in the state with $L = L_i$. If this time is sufficiently long, then it can be used as the time of observation of the states of the second system. Out of the entire time of observation of the second system (Tw_{L_i}) it spends a part equal to $(Tw_{L_i})w_{M_k}$ in the state with $M = M_k$. The probability that the first system is in the state with $L = L_i$ and that the second system is simultaneously in the state with $M = M_k$ is equal to

$$\lim_{T \rightarrow \infty} \frac{Tw_{L_i}w_{M_k}}{T} = w_{L_i}w_{M_k},$$

which accounts for the law of multiplication.

An important consequence of the law of addition of probabilities is the very obvious statement that the probability for a system to be in some allowed state is equal to unity. This means that we shall surely find our system in one of the states. The validity of the statement is seen from the fact that

$$\sum w_i = \sum \lim_{T \rightarrow \infty} \frac{t_i}{T} = \lim_{T \rightarrow \infty} \frac{\sum t_i}{T} = 1. \quad (2.10)$$

since, by definition, $T = \sum t_i$.

If the quantities characterizing the states of the system vary continuously, then instead of (2.10) we can write

$$\int dw = \int \rho(L) dL = 1. \quad (2.11)$$

In what follows we shall always assume probabilities to be normalized in such a way that the sum of all the probabilities is equal to unity. In this case we shall speak of the probability normalized to unity. In those cases when the initial probability distribution is not normalized, we shall always normalize it to unity.

§3. Mean values and fluctuations

It is now necessary to give a definition of the concept of the statistical mean value of a quantity which depends on the state of the system. The idea of the statistical mean will play a basic role in the subsequent discussion. The statistical mean is a natural generalization of the concept of the arithmetical mean which we are used to.

Suppose that we have a number of values of a certain quantity, for example the velocity of an arbitrary body. We understand the arithmetical mean to be the ratio of the sum of all these values to their total number, i.e. a sum of the form

$$\frac{\sum L_i N_i}{N},$$

where L_i is a value of L , N_i is the number of measurements leading to this value, and N is the total number of measurements.

By the statistical mean of a quantity L , which we shall denote by \bar{L} , we mean the limit of the ratio

$$\bar{L} = \lim_{N \rightarrow \infty} \frac{\sum L_i N_i}{N}.$$

Since $N_i = t_i/\Delta t$ and $N = T/\Delta t$, we can write

$$\bar{L} = \lim_{T \rightarrow \infty} \frac{\sum L_i t_i}{T} = \sum L_i w_{L_i}, \quad (3.1)$$

where t_i is the time during which the system is in the i th state, in which the quantity L has the value L_i , T is the total time of observation, and w_{L_i} is the probability that the quantity L has the value L_i . The summation is carried out over all states of the system. Formula (3.1) is the definition of the statistical mean. In what follows we shall for brevity omit the term "statistical" and say simply "mean value".

In the case of systems whose states vary continuously, so that instead of the probability w_i we must write dw , formula (3.1) should be rewritten in the form

$$\bar{L} = \int L dw = \int L \rho(L) dL, \quad (3.2)$$

where the integration is carried out over all possible states of the system.

In calculating mean values we shall make use of the following simple theorem: if there are two quantities L and M which are functions of the state, then the mean value of their sum $(\overline{L+M})$ is equal to the sum of the means $\bar{L} + \bar{M}$. For the proof we note that, by definition,

$$(\overline{L+M}) = \sum (L_i + M_i) w_i = \sum L_i w_i + \sum M_i w_i = \bar{L} + \bar{M}.$$

Assume that we know the distribution of the probabilities $w_{L,i}$ for a quantity L to take the values L_i . Then by means of formula (3.1) we can find the mean value of this quantity \bar{L} .

Thus, for example, knowing the probability distribution for different values of the energy of a system, one can calculate the mean value of the energy of this system. The question naturally arises: to what degree does the definition of the mean value characterize the real value of this quantity. In the example presented it can be asked to what degree the indication of the mean energy can characterize the actual energy of the system. It is clear that, if the deviations of the quantity from its mean value are sufficiently small, then the true value of the quantity can always, without much error, be replaced by its mean value.

In order to give a precise answer it is necessary to introduce a quantity which can characterize the deviation of the true value of a quantity L from its mean value \bar{L} .

At first sight it might seem that the difference $L - \bar{L}$ could be chosen as such a criterion. However, that is not so. The deviation of a quantity from its mean value can be large, but will nevertheless play a negligible role if it occurs relatively seldom. If, for example, considerable deviations of an energy from its mean value occur so seldom that the time interval between two successive deviations is very large in comparison with the observation time, then such deviations will not be noticed at all in the course of the time of observation. However, if deviations from the mean are not very large but occur often, then in this case the indication of merely the mean value \bar{L} does not characterize sufficiently the true value of the quantity L . One could try to choose as the criterion the mean value of the difference $L - \bar{L}$, i.e. $\overline{L - \bar{L}}$. However, this quantity is exactly equal to zero:

$$\overline{\Delta L} = \overline{L - \bar{L}} = \bar{L} - \bar{L} = 0.$$

(It should be noted that it is not necessary to carry out the secondary averaging of \bar{L} , since the mean \bar{L} is a constant quantity. But the mean value of a constant quantity is obviously equal to the quantity itself.)

The fact that the quantity $\overline{L - \bar{L}}$ is equal to zero expresses the fact that the deviations of L from \bar{L} in two directions, the direction of larger values and the direction of smaller values, occur equally frequently. In order that deviations from \bar{L} in the two directions may not cancel but be added, it is necessary to choose as the criterion not the mean difference $\overline{\Delta L} = \bar{L} - \bar{L}$, but the mean square of the difference $(\Delta L)^2$. The values of $(\Delta L)^2$ will be larger the larger the deviations of L from \bar{L} , independently of the signs of the devia-

tions, and the more often these deviations occur. The quantity $\overline{(\Delta L)^2} = \overline{(L - \bar{L})^2}$ is called the mean square deviation. The mean square deviation is an essentially positive quantity. It takes on its lowest possible value, zero, only in the case when L is always exactly equal to its mean value \bar{L} . Every deviation from the mean gives a contribution to the value of $\overline{(\Delta L)^2}$.

From the definition of $\overline{(\Delta L)^2}$ we have

$$\begin{aligned}\overline{(\Delta L)^2} &= \overline{(L - \bar{L})^2} = \overline{L^2 - 2L\bar{L} + (\bar{L})^2} \\ &= \bar{L}^2 - 2\bar{L}\bar{L} + (\bar{L})^2 = \bar{L}^2 - (\bar{L})^2.\end{aligned}\quad (3.3)$$

It is clear that for the absolute deviation to be small it is necessary that large departures of L from \bar{L} should have small probability, i.e. that they should occur relatively seldom. Thus, the quantity $\overline{(\Delta L)^2}$ can characterize the deviation of L from its mean value. If $\overline{(\Delta L)^2}$ is small, then the value of L is always close to its mean value. In this case the mean value \bar{L} can with sufficient accuracy characterize the value of L . The relative error which we shall incur by replacing L by its mean value \bar{L} can be estimated from the value of the quantity $\delta_L = \sqrt{\overline{(\Delta L)^2}}/\bar{L}$ called the relative fluctuation.

If $\delta_L \ll 1$, then this means that the value of L is on the average so close to \bar{L} that replacing of L by \bar{L} does not introduce any considerable error. We shall now prove a theorem of basic importance for subsequent developments. This theorem reads:

If there is a system consisting of N independent parts, then the relative fluctuation of any additive function* of the state, L , is inversely proportional to the square root of the number of parts N , i.e.

$$\delta_L = \frac{1}{\sqrt{N}}. \quad (3.4)$$

In order to prove this theorem we shall calculate the value of δ_L . By the definition of an additive quantity, $L = \sum_{k=1}^N L^{(k)}$, where $L^{(k)}$ is the value of the quantity L for the k th independent part of the system (in order to avoid misunderstandings, we write the index characterizing the number of the system as a superscript), and the summation is carried out over all independent parts constituting the system.

* By an additive function is meant a function possessing the property that the value of this function for a complex system is equal to the sum of its values for all independent parts.

From the law of addition of probabilities it follows that

$$\bar{L} = \sum_{k=1}^N \overline{L^{(k)}}. \quad (3.5)$$

We now calculate the mean square deviation of L , i.e. the quantity

$$\overline{(\Delta L)^2} = \overline{\left[\Delta \sum_{k=1}^N L^{(k)} \right]^2}.$$

For simplicity we assume at first that the system consists of only two independent parts. Then we have

$$\overline{[\Delta(L_1 + L_2)]^2} = \overline{(\Delta L_1)^2} + \overline{2\Delta L_1 \cdot \Delta L_2} + \overline{(\Delta L_2)^2}.$$

Now, since L_1 and L_2 are independent quantities, the mean of the product $(\Delta L_1)(\Delta L_2)$ is equal to the product of the means

$$\overline{(\Delta L_1)(\Delta L_2)} = \overline{(\Delta L_1)} \cdot \overline{(\Delta L_2)}.$$

But $\overline{(\Delta L_1)} = \overline{(\Delta L_2)} = 0$, so that

$$\overline{[\Delta(L_1 + L_2)]^2} = \overline{(\Delta L_1)^2} + \overline{(\Delta L_2)^2},$$

i.e. the mean square deviation of a system of two independent quantities is equal to the sum of the mean square deviations of these quantities. Generalizing this to the case of N independent parts constituting a system we can write

$$\overline{\left[\Delta \left(\sum_{k=1}^N L^{(k)} \right) \right]^2} = \sum (\Delta L^{(k)})^2. \quad (3.6)$$

The number of terms in the sum (3.6) is equal to the number of independent parts in the system, i.e. to N . We assume that the fluctuations in different independent parts of the system are of the same order of magnitude (since all the parts of the system are equivalent). Then the value of the sum on the right-hand side of formula (3.6) is proportional to the number of terms, i.e. to the value of N , so that

$$\left[\Delta \left(\sum_{k=1}^N L^{(k)} \right) \right]^2 \sim N. \quad (3.7)$$

The mean value \bar{L} is also proportional to the number of terms in the sum of formula (3.6), i.e. to N . Hence the relative fluctuation of the quantity L is equal to

$$\delta_L = \frac{\sqrt{(\Delta L)^2}}{\bar{L}} = \frac{\sqrt{\left[\Delta \left(\sum L^{(k)} \right) \right]^2}}{\sum \bar{L}^{(k)}} \sim \frac{\sqrt{N}}{N} \sim \frac{1}{\sqrt{N}}. \quad (3.8)$$

Thus, the theorem is proved.

As has already been pointed out in the introduction, the problem of statistical physics is the study of the properties of macroscopic systems consisting of an enormous number of particles — atoms or molecules. We shall see later that the methods of investigating the properties of such systems are based on the application of statistical laws. The application of these laws allows one to find the mean values of different quantities characterizing the state of a system. From the theorem presented above it follows that the relative fluctuations of all physical quantities whose value for the entire system is equal to the sum of their values for all the particles are inversely proportional to the square root of the number of particles. Since the number of particles in a macroscopic system is expressed as a rule by enormous numbers (of the order of 6×10^{23}), the relative fluctuation of any additive quantity turns out to be practically equal to zero. This means that all additive quantities have values very close to the means. Therefore the replacement of actual quantities by their mean values can be carried out with a high degree of accuracy. The mean values of different quantities, calculated on the basis of the laws of statistical physics, agree to a very high degree of accuracy with their true values. This means that predictions based on probability in practice assume a completely reliable character.

Imagine, for example, that we want to find the pressure exerted by a mole of a gas on the walls of the container in which it is confined. By means of the propositions of statistical physics it turns out that it is possible to calculate the mean pressure \bar{p} of the gas. The true pressure p exerted on the wall is not equal to the mean pressure. Depending on the complex laws of motion of molecules in a gas, it will take on different values varying rapidly in time (fig. III.4), which can be both larger and smaller than the mean pressure.

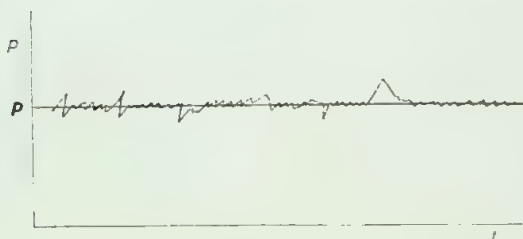


Fig. III.4

Nevertheless, the theorem on fluctuations shows that the relative error which we shall incur by replacing the true pressure varying in time by its mean value (shown in fig. III.4 by the horizontal line) will be of the order of $\delta_p \approx (6 \times 10^{23})^{-\frac{1}{2}}$, i.e. the error will amount to $\approx 10^{-10}\%$.

It is obvious that such an error lies far beyond the limits of accuracy of measurements by the best manometers, and is of no practical importance. Hence we can make use of the mean value of pressure, without being concerned that any error will be committed. The same holds also for other functions of the state of the system. Examples of these functions will be given later.

§4. Normal distribution and moments

Returning to the discussion of the properties of the additive quantity $L = \sum_{k=1}^N L^{(k)}$ it should be noted that in the theory of probability the following very important theorem, called the central limit theorem, has been proved: "as the number of terms in the sum increases (as $N \rightarrow \infty$) the statistical probability distribution for the quantity L tends to the normal (Gaussian) distribution having the form:

$$\rho(L)dL = \frac{1}{\sqrt{2\pi(\Delta L)^2}} \exp \left[-\frac{(L-\bar{L})^2}{2(\Delta L)^2} \right] dL. \quad (4.1)$$

Applied to physical systems this means that the normal distribution for additive quantities, for example energy, will be established in any physical system containing a sufficiently large number of independent particles. We shall not present the proof of the central limit theorem, but shall confine ourselves to the consideration of a characteristic example.

Consider a system of N identical statistically independent particles. Let the probability for one of the particles to get into the p th state be equal to p . We shall find the probability that n particles be found in this state. For this we write the probability that n particles be found in the state p and that the remaining $(N-n)$ particles be found in other states in the form $p^n(1-p)^{N-n}$ (on the basis of (2.9), since the particles are independent).

The number of ways in which n arbitrary particles can be chosen out of the total number of particles N is equal to the number of combinations of N elements taken n at a time. The latter is equal to

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}.$$

Hence the total probability that n arbitrarily chosen particles be found simultaneously in the p th state is equal to

$$w_N(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}. \quad (4.2)$$

The expression obtained is called the binomial law. We shall now assume that the system contains a very large number of particles, so that $N \gg n$. Then omitting n in the exponent of the last factor, we can write

$$\begin{aligned} w_N(n) &\cong \frac{N!}{n!(N-n)!} p^n (1-p)^N \\ &= \frac{N(N-1) \dots (N-n+1)}{n!} p^n (1-p)^N = \frac{(\bar{n})^n}{n!} \left(1 - \frac{\bar{n}}{N}\right)^N, \end{aligned}$$

where $\bar{n} = pN$. It is obvious that \bar{n} represents the mean number of particles in the p th state. In the limit $N \rightarrow \infty$ we obtain

$$w(n) = \lim_{N \rightarrow \infty} w_N(n) = \frac{(\bar{n})^n e^{-\bar{n}}}{n!}. \quad (4.3)$$

This formula is called the Poisson formula. Finally, we shall find the asymptotic expression of the Poisson formula for the case when not only is N very large but also the number of particles n in a given state is large. This means that n and \bar{n} can be considered to be large numbers (in comparison with unity), and the difference $n - \bar{n} \ll \bar{n}$.

Taking the logarithm of the Poisson formula, we have

$$\ln w(n) = n \ln \bar{n} - \bar{n} - \ln n! .$$

Making use of Stirling's formula (Appendix IV), we can write

$$\ln n! \cong n \ln n - n ,$$

so that

$$\begin{aligned} \ln w(n) &= n \ln (\bar{n}/n) + (n - \bar{n}) \\ &= -n \left\{ \ln \left(\frac{n}{\bar{n}} \right) + \frac{n - \bar{n}}{n} \right\} . \end{aligned}$$

Using $\ln(1+x) = x - \frac{1}{2}x^2 \dots$ and $\bar{n} \sim n$, we obtain

$$w(n) = \text{const} \exp \left[-\frac{(n - \bar{n})^2}{2\bar{n}} \right] .$$

The constant is found from the normalization condition. For large values of n the summation can be replaced by integration. We then obtain

$$dw(n) = w(n)dn = \frac{1}{\sqrt{2\pi\bar{n}}} \exp \left[-\frac{(n - \bar{n})^2}{2\bar{n}} \right] dn , \quad (4.4)$$

i.e. the normal (Gaussian) probability distribution. The mean square deviation of the number of particles in the state considered is equal to

$$\overline{(\Delta n)^2} = \overline{(n - \bar{n})^2} = \frac{1}{\sqrt{2\pi\bar{n}}} \int (n - \bar{n})^2 dw(n) = \bar{n} . \quad (4.5)$$

In the case given formula (3.7) turns out to be exact, not approximate. Hence the Gaussian distribution can be written in the form

$$dw(n) = \rho(n)dn = \frac{1}{\sqrt{2\pi(\Delta n)^2}} \exp \left[-\frac{(n - \bar{n})^2}{2(\Delta n)^2} \right] dn , \quad (4.6)$$

which is the same as (4.1). We see, in this particular example, that for large values of the numbers N and n the normal probability distribution is established, since the deviation of the numbers n from their mean values are on the average sufficiently small.

The mean square deviation characterizes the effective width of the normal distribution. The smaller $(\Delta n)^2$, or in the general case $(\Delta L)^2$, the smaller the width of the Gaussian distribution. In the limit $(\Delta L)^2 \rightarrow 0$ the Gaussian distribution goes over into the δ -function. In this case the probability of finding a value $L \neq \bar{L}$ tends to zero, while the probability of the value $L = \bar{L}$ tends to unity. The normal distribution has a symmetric character, so that $w(L) = w(-L)$, i.e. the probability of a deviation from the mean is the same in both directions. If a distribution is not Gaussian, then, generally speaking, it is not symmetric with respect to the sign of L . The degree of asymmetry of the distribution is characterized by a quantity called the skewness and equal to

$$(\overline{\Delta L})^3 = \bar{L}^3 - 3\bar{L}(\overline{\Delta L})^2 = \bar{L}^3 - 3\bar{L}[\bar{L}^2 - (\bar{L})^2]. \quad (4.7)$$

The mean square deviation and the skewness are expressed in terms of quantities \bar{L} and \bar{L}^n ($n=2,3$). The latter, defined in general form by the formula

$$\bar{L}^n = \int L^n \rho(L) dL, \quad (4.8)$$

are called the moments of n th order. It turns out that, if $\rho(L)$ is an analytical function having derivatives of all orders, then the set of moments of all orders $(\bar{L}, \bar{L}^2, \bar{L}^3, \dots)$ completely determines the form of the function $\rho(L)$. Indeed, carrying out the Fourier transformation on $\rho(L)$, we have

$$\psi(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \rho(L) e^{i\omega L} dL. \quad (4.9)$$

The function $\psi(\omega)$ is called the characteristic function of the probability distribution.

Differentiating (4.9) with respect to ω and setting $\omega = 0$, we find

$$\begin{aligned} \left(\frac{d\psi}{d\omega} \right)_{\omega=0} &= \frac{i}{2\pi} \int_{-\infty}^{\infty} L \rho(L) dL = \left(\frac{i}{2\pi} \right) \bar{L}, \\ \left(\frac{d^2\psi}{d\omega^2} \right)_{\omega=0} &= \frac{i^2}{2\pi} \int_{-\infty}^{\infty} L^2 \rho(L) dL = \left(\frac{i^2}{2\pi} \right) \bar{L}^2, \end{aligned}$$

$$\dots\dots\dots,$$

$$\left(\frac{d^n \psi}{d\omega^n}\right)_{\omega=0} = \frac{i^n}{2\pi} \int_{-\infty}^{\infty} L^n \rho(L) dL = \left(\frac{i^n}{2\pi}\right) \overline{L^n}.$$

Hence, if all moments are known, then the coefficients in the expression

$$\psi(\omega) = \psi(0) + \psi'_0 \frac{\omega}{1!} + \psi''_0 \frac{\omega^2}{2!} + \dots \quad (4.10)$$

are known and, consequently, the function $\psi(\omega)$ itself is known. Then the probability distribution is obtained directly from the transformation

$$\rho(L) = \int_{-\infty}^{\infty} \psi(\omega) e^{-i\omega L} d\omega. \quad (4.11)$$

Sometimes the moments $\overline{L^n}$ or, in any case, the first few moments are known, but the probability distribution itself is not known. Then, by finding exactly the characteristic function $\psi(\omega)$ (or approximately), one can find exactly (or approximately) the probability distribution $\rho(L)$.

§5. The correlation function

In what follows we shall often have to consider random functions. A random function is a function $f(x)$ the values of which do not depend in an unambiguous way on the variable x . For a fixed value of x , the function $f(x)$ can take on randomly all possible values. In this case one can only speak of the probability that, for a given x , the function $f(x)$ has a value lying between $f(x)$ and $f(x) + df(x)$. For concreteness we shall assume in what follows that the random quantity depends on time, i.e. we shall consider a random function of time $f(t)$. A process described by a random function of time is called a stochastic process. Physical examples of stochastic processes and random functions depending on time will be given below.

The most important quantitative characteristic of random processes is their correlation function (or, more precisely, autocorrelation). By the correlation function $K(\tau)$ is meant the mean value (with respect to time or with respect to the ensemble) of the product of a random function describing a stochastic process in a certain system, taken at time t , and the same function taken at time $t + \tau$:

$$K(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(t) f(t+\tau) dt, \quad (5.1)$$

where τ can be positive as well as negative. For brevity we shall write

$$K(\tau) = \overline{f(t) f(t+\tau)}. \quad (5.2)$$

The bar denotes averaging over time.

In addition to averaging over time, averaging with respect to an ensemble of identical physical systems in which a random physical process takes place, can be carried out. By virtue of what was said in §2, the two averages are equivalent. Hence we can write

$$K(\tau) = \langle f(t) f(t+\tau) \rangle. \quad (5.3)$$

The brackets $\langle \rangle$ denote the average with respect to the ensemble.

The correlation function $K(\tau)$ is a quantitative measure of the connection between the values of a random function at successive instants of time. In other words, $K(\tau)$ is a measure of the rate of change in time of a function $f(t)$ describing a stochastic process.

The values of the correlation function depend only on τ , and not on the choice of the value of t . Indeed, because of the uniformity of time, a change of the origin cannot affect mean values, so that

$$K(\tau) = \overline{f(t) f(t+\tau)} = \overline{f(t') f(t'+\tau)}. \quad (5.4)$$

If the values of a random function $f(t)$ vary so rapidly that its value at an instant $t + \tau$ does not depend at all on the value at the instant t , then

$$K(\tau) = \overline{f(t)} \overline{f(t+\tau)} = 0. \quad (5.5)$$

When $\tau \rightarrow \infty$ we have the obvious equality

$$K(\tau \rightarrow \infty) \rightarrow 0, \quad (5.6)$$

called the property of attenuation of correlation at infinity.

In the other limiting case $\tau = 0$ the equality

$$K(0) = [f(t)]^2 = \langle f^2(t) \rangle \quad (5.7)$$

shows that $K(0)$ is the same as the mean-square value (the second moment) of the random function $f(t)$. Finally, the symmetry of the stochastic process in time allows one to write the condition

$$K(\tau) = K(-\tau) . \quad (5.8)$$

The actual form of a correlation function depends, of course, on the nature of the random process. However, there exists a theorem relating two important characteristics of a random process – the correlation function and the so-called spectral density. The random function $f(t)$ can be expanded in a Fourier integral:

$$f(t) = \int_{-\infty}^{\infty} e^{i\omega t} f(\omega) d\omega . \quad (5.9)$$

The frequencies ω form a continuous spectrum.

We write the mean-square value $\langle f^2 \rangle$ in the form

$$\langle f^2(t) \rangle = \int_{-\infty}^{\infty} I(\omega) d\omega = 2 \int_0^{\infty} I(\omega) d\omega , \quad (5.10)$$

where the function $I(\omega)$ is called the spectral density. By definition $I(\omega)$ is an essentially positive function, $I(\omega)$ being equal to $I(-\omega)$.

Substituting (5.9) into the definition (5.7) of $\langle f^2 \rangle$, we find

$$\langle f^2(t) \rangle = \iint d\omega d\omega' e^{i(\omega+\omega')t} \langle f(\omega) f(\omega') \rangle . \quad (5.11)$$

We write the expression for $\langle f(\omega) f(\omega') \rangle$, making use of the inversion formula:

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt . \quad (5.12)$$

We then obtain

$$\langle f(\omega) f(\omega') \rangle = \frac{1}{4\pi^2} \iint \langle f(t) f(t') \rangle e^{i(\omega t + \omega' t')} dt dt' .$$

Assuming $t' = t + \tau$, we have

$$\begin{aligned}\langle f(\omega) f(\omega') \rangle &= \frac{1}{4\pi^2} \iint e^{i(\omega+\omega')t} e^{i\omega'\tau} \langle f(t) f(t+\tau) \rangle dt d\tau \\ &= \frac{1}{4\pi^2} \iint e^{i(\omega+\omega')t} e^{i\omega'\tau} K(\tau) dt d\tau.\end{aligned}\quad (5.13)$$

According to (5.4) the correlation function $K(\tau)$ does not depend on t , and in the last expression one can take $K(\tau)$ out of the integral with respect to t :

$$\begin{aligned}\langle f(\omega) f(\omega') \rangle &= \frac{1}{4\pi^2} \int e^{i\omega'\tau} K(\tau) d\tau \int_{-\infty}^{\infty} e^{i(\omega+\omega')t} dt \\ &= \frac{1}{2\pi} \int e^{i\omega'\tau} K(\tau) d\tau \delta(\omega+\omega').\end{aligned}\quad (5.14)$$

Substituting (5.14) into (5.11), we have

$$\begin{aligned}\langle f^2(t) \rangle &= \frac{1}{2\pi} \int d\tau d\omega d\omega' e^{i\omega'\tau} K(\tau) \delta(\omega+\omega') \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int K(\tau) e^{-i\omega\tau} d\omega d\tau.\end{aligned}\quad (5.15)$$

Comparing (5.15) and (5.10), we finally obtain

$$I(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega\tau} K(\tau) d\tau.\quad (5.16)$$

Inverting the last integral, we can also write

$$K(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} I(\omega) d\omega.\quad (5.17)$$

Formulae (5.17) and (5.16) represent the content of the Wiener-Khintchine theorem. They relate the correlation function and the spectral density. The latter turns out to be the Fourier transform of the correlation function.

In studying stochastic processes in physical systems we shall encounter applications of the Wiener-Khintchine theorem.

The Kinetic Theory of Gases

§6. The simplest statistical system — an ideal gas

It is natural to begin the study of systems containing a very large number of particles with the simplest case — an ideal gas.

The density of matter in the gaseous state is small, so that the mean distance between particles is very large in comparison with the geometrical dimensions of the particles, the atoms or molecules. Because of this each of the particles is relatively distant from the other gas particles for most of the time of motion.

The forces of intermolecular interaction decrease rapidly with increasing distance and become negligibly small when molecules are at a distance considerably exceeding their geometrical size. Thus, a characteristic feature of the motion of molecules in a gas is the smallness of the intermolecular interaction during most of the time. Because of the absence of interaction the gas molecules move rectilinearly and uniformly until a collision takes place between a given molecule and any other molecule or the wall of the container. In collisions of gas molecules with each other or with the molecules of the wall of the container the molecules can be considered to be undeformable. This means that the collisions between molecules obey the same laws as the collisions between ordinary hard balls. In the process of a collision an exchange

of kinetic energy and momentum takes place between molecules. Analogously, when a gas molecule collides with the wall of the container or, more precisely, with a molecule of the substance of this wall, it can be assumed that the gas molecule is reflected elastically from the wall.

An ideal gas is a statistical system the particles of which interact with one another only through collisions, while during the remaining time they move as free particles.

The motion of every gas molecule is determined strictly by the laws of mechanics (in the first approximation those of classical mechanics). Hence, by integrating the equations of motion of all the molecules constituting a gas, one could in principle find the trajectory of every one of the molecules. However, such a calculation would, in practice, present great difficulties. Even the integration of the equations of motion of three interacting material points (the three-body problem) is a very complicated problem, the general case being still unsolved. The general solution of the four-body problem is so complicated that ways of solving it are not even known. And in a gas the number of interacting particles is expressed by numbers of the order of 10^{20} . From the macroscopic point of view, innumerable collisions between the molecules and with the wall of the container take place in an extremely short time interval. Hence, in order to find the trajectories of all the molecules of a gas it would be necessary to write down and solve 3×10^{20} inter-related equations of motion taking into account the corresponding initial conditions. From this it is clear that such a problem is difficult not only in practice but also in principle.

At first sight it may seem that the above statement deprives us of any possibility of studying physical laws in systems consisting of a very large number of particles. In reality, however, this is not so. Although every one of the particles constituting a system is in itself a "mechanical system" and obeys the laws of mechanics, the set of an enormous number of molecules is a system differing qualitatively from a system consisting of a small number of particles. Laws of a special type, which are not at all inherent in simple mechanical systems, and which are called statistical laws, are shown by it.

Consider a gas consisting of an enormous number of molecules confined in a closed container. Such a gas represents a mechanical system with an enormous number of degrees of freedom. Knowing the initial conditions, it would, in principle, be possible to integrate the equations of motion of all the gas molecules and to find their trajectories. Putting aside the question as to the practical feasibility of such a calculation, it should be noted that such a solution would be of no interest. We know from experience that the properties of a gas do not depend at all on the initial conditions, i.e., the initial

positions and velocities of the molecules. Thus, for example, the properties of a gas in a closed container in no way depend on the method of filling the container: independently of whether the gas flowed into the container through one opening and slowly, or through two openings and rapidly, after the lapse of a certain time interval from the admission, the gas will come into a definite state in which it will henceforth be found.

It is said that the gas comes into an equilibrium state. The properties of a gas in an equilibrium state do not depend on its previous history and do not change in time.

It is well known from experience that a gas always tends to occupy the entire volume available completely and uniformly. Hence a form of motion of a gas in which the density is equal in different parts of the container is ruled out or, more precisely, is extremely improbable. No obvious contradiction with the laws of mechanics would appear in this case. If the state of a system depended on the initial conditions, then the latter could in principle be chosen in such a way that the density of the gas in different parts of the container would be different. The fact that the state of a gas does not depend on the initial conditions of the positions and velocities of its molecules makes of no use the knowledge of the trajectories of individual molecules. Assume that we succeeded in overcoming all mathematical difficulties and found the trajectory of an individual molecule. Suppose it turned out that the trajectory of the given molecule lay almost entirely in one of the corners of a cube. It is clear that we could not draw from this any conclusion about the behaviour of the entire gas.

A gas as a whole, containing an enormous number of particles, appears to be a system differing qualitatively from an individual molecule, and its behaviour obeys different, statistical, laws. In this connection one of the founders of statistical physics, Smoluchowski, wrote that even if we were able to find the trajectories of the gas molecules we would still make use of laws of probability theory in describing the properties of the gas.

In finding statistical laws we shall look for the mean value of the quantities characterizing the state of a gas as a whole. Because the number of particles in a gas is very large, it follows from the results of §5 that the mean values found will agree to a high degree of accuracy with the true values of these quantities.

Earlier reasoning allows us to bring forward the most important feature of "mass" processes, i.e. processes which are characterized by the presence of a large number of more or less equivalent events. This feature is the fact that in such processes peculiar statistical laws are shown which are not inherent in individual systems or processes.

Consider a closed container filled by a gas. Assume that an equilibrium state is established in the gas. We shall try to find the statistical laws determining the behaviour of the gas. In correspondence with direct experimental data we shall assume that the molecules of the gas are distributed over the whole volume of the closed container with a uniform density (i.e. that the number of molecules per unit volume is constant everywhere in container). We shall also assume that the molecules of the gas have velocities which are uniformly distributed over all directions in space. This means that the number of molecules moving in any direction must be the same. If this were not so and if there existed a preferential direction of motion of molecules, then a flow of gas would arise in this direction. It follows from experiment that in a gas confined in a closed container and not subjected to any action from without no stationary gas flow can arise. The assumption of a uniform distribution of molecules in space and a uniform velocity distribution over all directions is called the assumption of molecular disorder.

The question naturally arises as to how the uniform distribution of the velocities of molecules in all directions is established. It is clear that, if molecules did not interact with one another at all, then there would be no way of changing the initial direction of motion of a molecule. Hence the presence or absence of a directed flow would entirely be determined by the initial conditions. The establishment of molecular disorder is due to the existence of interaction between the molecules. Because of collisions between molecules their directions of motion change continually, and a random motion with a uniform velocity distribution over directions in space is established in the gas.

The role of molecular collisions does not only give rise to the establishment of a uniform velocity distribution over directions. In molecular collisions a change in the absolute value of the velocities of the molecules takes place in addition to a change of the direction of flight. If initially all molecules had the same velocity, then random collisions between them would lead to the fact that some of the molecules would randomly get an excess of kinetic energy at the expense of other molecules which would correspondingly lose a part of their energy. Owing to this the equality of the velocities of gas molecules would be violated, and a certain number of molecules having larger and smaller velocities would appear in the gas. In other words, a certain distribution of molecules over velocities will arise in the gas. A number of molecules having large velocities, and a number of molecules with medium and small velocities will appear.

Our problem is to find the velocity distribution of the molecules of an ideal gas. This distribution will be characterized by the mean number of molecules having a given value of the velocity.

From the assumption of the random character of molecular motion it follows that the appearance of molecules with a variety of velocities is possible, so that the distribution of molecules can be characterized by a certain continuous function. Since the velocities of motion of molecules vary continually, one has, of course, to speak not of the number of molecules having a given exact velocity but of the number of molecules having a velocity close to the given one.

§ 7. The Maxwell distribution

Let us denote by dn_v the mean number of molecules in unit volume of a gas having velocity components lying within the interval between v_x and $v_x + dv_x$, v_y and $v_y + dv_y$, v_z and $v_z + dv_z$.

We shall assume that the gas is in a stationary state, so that the state of molecular disorder, which does not change in time, is established. The number of particles with given velocity components does not depend on time.

It is clear that the mean number of molecules dn_v can be written in the following form

$$dn_v = n(v_x, v_y, v_z) dv = n(\mathbf{v}) dv_x dv_y dv_z, \quad (7.1)$$

where $n(v_x, v_y, v_z) = n(\mathbf{v})$ is the mean number of molecules with velocity components v_x, v_y, v_z , in unit interval. The function $n(\mathbf{v})$ is called the velocity distribution function of the molecules.

Since all directions of motion of molecules in space are equivalent, the velocity distribution must be isotropic and the distribution function $n(\mathbf{v})$ cannot depend on the direction of the velocity. This means that $n(v_x, v_y, v_z)$ cannot be an arbitrary function of the velocity components v_x, v_y, v_z but must be a function of the argument $v = |\mathbf{v}| = (v_x^2 + v_y^2 + v_z^2)^{1/2}$ i.e. of the absolute value of the velocity, and

$$dn_v = dn_v. \quad (7.2)$$

Changing from velocity components to the absolute value of the velocity and its direction characterized by the polar angles ϑ and ψ , we can, by virtue of (1.67) write that

$$dn_v = n(v) v^2 dv \sin \vartheta d\vartheta d\psi. \quad (7.3)$$

The total number of particles in unit volume $n = N/V$ determines the normalization condition

$$n = \int n(v) v^2 dv \sin \vartheta d\vartheta d\psi = 4\pi \int n(v) v^2 dv. \quad (7.4)$$

The range of integration with respect to the velocity components v_x, v_y, v_z or with respect to the absolute value of the velocity v will be considered later.

Our problem is to find the explicit form of the distribution function $n(v)$. Here we shall confine ourselves to the simplest, although not completely rigorous, derivation of a form of the distribution function called the Maxwell distribution function or, briefly, the Maxwell distribution. In this derivation the role of molecular collisions and the proposition of molecular disorder, in the establishment of the equilibrium velocity distribution of molecules, is particularly clear.

Let us consider the process of collision between two particles moving with velocities \mathbf{v}_1 and \mathbf{v}_2 . Since the forces of intermolecular interaction decrease rapidly with increasing distance and are, in fact, different from zero only at the moment of direct contact of the particles, we can replace the real process of collision by an idealized scheme of the elastic collision between two material points. Here the concrete form of the interaction forces does not play any role. As a result of the collision, let the velocities of the molecules be changed. After the collision the first particle moves with a velocity \mathbf{v}_3 , and the second one with a velocity \mathbf{v}_4 . The number of such collisions per unit time in unit volume of the gas must be proportional to the number of molecules with velocity \mathbf{v}_1 and the number with velocity \mathbf{v}_2 , i.e. to the product $n(v_1) n(v_2)$. Consider further the reverse process of collision. Then the velocities of the molecules change from the values \mathbf{v}_3 and \mathbf{v}_4 to the values \mathbf{v}_1 and \mathbf{v}_2 . The number of such collisions per unit time in unit volume is proportional to the number of molecules with velocity \mathbf{v}_3 and the number with velocity \mathbf{v}_4 , i.e. to $n(v_3) n(v_4)$.

Because of our assumption that the number of molecules with given values of velocity does not change by the processes of molecular collision in a gas in a stationary state, we can consider that the number of molecules for which the velocities are changing from the values $\mathbf{v}_1, \mathbf{v}_2$ to the values $\mathbf{v}_3, \mathbf{v}_4$ is equal to the number of molecules for which the velocities are changing from $\mathbf{v}_3, \mathbf{v}_4$ to $\mathbf{v}_1, \mathbf{v}_2$, i.e. it can be assumed that

$$n(v_1) n(v_2) = n(v_3) n(v_4). \quad (7.5)$$

The equality (7.5) expresses the balance of particles acquiring and losing corresponding velocities.

Since in the process of collision the energy of the molecules is conserved, for the direct and reverse process we can write

$$v_1^2 + v_2^2 = v_3^2 + v_4^2. \quad (7.6)$$

The equality (7.6) expresses the energy conservation law for the collision. The common factor $\frac{1}{2}m$, figuring in both sides, has here been dropped.

The equalities (7.4), (7.5) and (7.6) represent the whole set of conditions which must be satisfied by the distribution function sought.

From formula (7.6), into which only the squares of the velocities enter, it is seen that the functional eq. (7.5) will look much simpler if the square of the velocity is chosen instead of its absolute value as the argument of the distribution function writing the required function in the form $n(v^2)$. This does not change the essence of the problem, but allows one to rewrite (7.5) in a form which is simpler from the mathematical point of view

$$n(v_1^2) n(v_2^2) = n(v_3^2) n(v_4^2). \quad (7.7)$$

The functional eq. (7.7) is easy to transform into a simple differential equation. For this we express v_4 in terms of v_1, v_2 and v_3 by means of (7.6) and rewrite (7.7) in the form

$$n(v_1^2) n(v_2^2) = n(v_3^2) n(v_1^2 + v_2^2 - v_3^2).$$

Taking the logarithm of this equality, we have

$$\ln n(v_1^2) + \ln n(v_2^2) = \ln n(v_3^2) + \ln n(v_1^2 + v_2^2 - v_3^2). \quad (7.8)$$

We differentiate the above equality with respect to the argument v_1^2 . Here we have to keep in mind that eq. (7.8) is valid for completely arbitrary and independent values of v_1, v_2 and v_3 [the value of v_4 is determined by formula (7.6)]. We have

$$\frac{1}{n(v_1^2)} \frac{dn(v_1^2)}{dv_1^2} = \frac{1}{n(v_1^2 + v_2^2 - v_3^2)} \frac{dn(v_1^2 + v_2^2 - v_3^2)}{d(v_1^2 + v_2^2 - v_3^2)}.$$

Analogously,

$$\frac{1}{n(v_2^2)} \frac{dn(v_2^2)}{dv_2^2} = \frac{1}{n(v_1^2 + v_2^2 - v_3^2)} \frac{dn(v_1^2 + v_2^2 - v_3^2)}{d(v_1^2 + v_2^2 - v_3^2)}.$$

Comparing the two equations, we obtain

$$\frac{1}{n(v_1^2)} \frac{dn(v_1^2)}{dv_1^2} = \frac{1}{n(v_2^2)} \frac{dn(v_2^2)}{dv_2^2}. \quad (7.9)$$

Since v_1 and v_2 are independent variables, and the equality (7.9) must hold for completely arbitrary values of the independent quantities v_1 and v_2 , it is clear that it can be fulfilled only when the right and left hand sides of (7.9) are each equal to a constant.

We denote this constant by $-\alpha$. Then instead of (7.9) we can write

$$\frac{1}{n(v^2)} \frac{dn(v^2)}{dv^2} = -\alpha. \quad (7.10)$$

In the above equation we have dropped the subscript on the velocity, since from the preceding considerations it is clear that the equation must be valid for any value of the velocity.

Integrating (7.10) we find

$$n(v^2) = A e^{-\alpha v^2}, \quad (7.11)$$

where A is the integration constant. The integration constant A can be determined from the normalization condition (7.4). By virtue of (7.11) and (7.4) we have

$$4\pi A \int e^{-\alpha v^2} v^2 dv = n. \quad (7.12)$$

Now let us consider the integration range in the normalization condition (7.12). The lower limit of integration in (7.12) corresponds to the lowest possible value of the velocity v which is, obviously, equal to zero. As to the upper limit, we cannot, of course, indicate the value of the highest velocity which can be possessed by a molecule of the gas. However, the form of the distribution entering into (7.12) shows that, in essence, there is no need to know this value. The integrand decreases so rapidly with increasing argument that we shall not make any error by substituting infinity for the upper limit in (7.12). Hence the normalization condition (7.12) can be written in the form

$$4\pi A \int_0^{\infty} e^{-\alpha v^2} v^2 dv = n. \quad (7.13)$$

From (7.13) it follows directly that $\alpha > 0$. Otherwise the integral does not exist. The integration gives

$$A = n(\alpha/\pi)^{\frac{3}{2}}. \quad (7.14)$$

Finally, the distribution function can be written in the form

$$n(v) = n(\alpha/\pi)^{\frac{3}{2}} e^{-\alpha v^2}. \quad (7.15)$$

The number of molecules per unit volume with velocities between v and $v+dv$ is, thus, equal to

$$dn_v = 4\pi n(\alpha/\pi)^{\frac{3}{2}} e^{-\alpha v^2} v^2 dv. \quad (7.16)$$

Formula (7.16) is called the Maxwell distribution.

In addition to the velocity distribution one can also write the distribution over the components of velocities

$$dn_v = n(\alpha/\pi)^{\frac{3}{2}} \exp[-\alpha(v_x^2 + v_y^2 + v_z^2)] dv_x dv_y dv_z. \quad (7.17)$$

The transition from (7.16) to (7.17) corresponds to the ordinary transformation of polar coordinates into Cartesian coordinates.

In formula (7.17) it can be assumed that the components of the velocity vary from $-\infty$ to $+\infty$.

It should be noted that the velocity component distribution function can be written in the form of the product of three velocity component distribution functions

$$\begin{aligned} dn_v &= dn_{v_x} dn_{v_y} dn_{v_z} = \\ &= n(\sqrt{\alpha/\pi} e^{-\alpha v_x^2} dv_x) (\sqrt{\alpha/\pi} e^{-\alpha v_y^2} dv_y) (\sqrt{\alpha/\pi} e^{-\alpha v_z^2} dv_z). \end{aligned} \quad (7.18)$$

Before proceeding to a discussion of the results following from the distributions (7.16) and (7.17) it is necessary to explain the meaning of the parameter α appearing in them.

§8. Collisions of molecules with the wall of the container. Pressure. The connection of the parameter α with the absolute temperature

During their motion the molecules of a gas confined in a container undergo

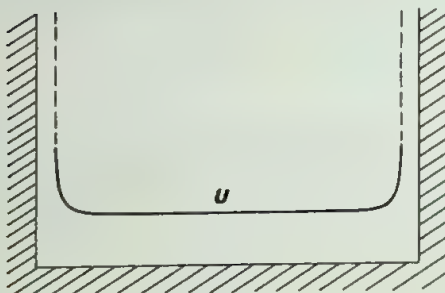


Fig. III.5

collisions with its walls. The walls of the container have molecular structure and do not form geometrically sharp boundaries. A gas molecule approaching the wall undergoes a very strong repulsion by the wall molecules, and is reflected back into the container. Fig. III.5 shows schematically the form of the potential energy of a molecule near the walls of the container. The latter can be considered as an infinitely high impenetrable potential barrier for molecules. It can be assumed that the reflection of a molecule from the wall of the container is completely elastic. This means that, in the reflection, the velocity component perpendicular to the surface of the wall changes to its exact opposite.

Consider a certain wall area dS , perpendicular to the x -axis. Then a molecule having velocity components v_x, v_y, v_z acquires, when reflected, velocity components $-v_x, v_y, v_z$. In the reflection of the molecule from the area the momentum component along the x -axis changes from the value mv_x to $-mv_x$, i.e. by the amount $2mv_x$. This momentum is transferred to the reflecting wall.

Thus, collisions of molecules with the wall lead to the appearance of a force acting on the surface of the container. We identify the force acting on unit surface of the wall by all molecules of the gas with the macroscopic pressure. This statement which, in essence, is the basis of the kinetic theory of gases, seemed at one time to be very radical. However, at present it appears to be natural and completely obvious.

In order to find the pressure exerted on the wall it is necessary to calculate the total change in the momentum of gas molecules undergoing reflection from unit surface area of the container in unit time. It is obviously equal to the change in the momentum in one collision with the wall multiplied by the total number of collisions per cm^2 of the surface per sec. The change in the momentum is equal to $2mv_x$. Multiplying this expression by the number of

impacts per cm^2 per sec by molecules having a given velocity component v_x , and summing or, more exactly, integrating this product over all values of v_x , we find the pressure sought. In unit time the surface of the wall will be reached by all molecules which are a distance from it smaller than or equal to v_x (since v_x is the path traversed in unit time by a molecule moving in the positive direction of the x -axis). All molecules contained in a parallelepiped of a height v_x and a base of 1 cm^2 will impinge on 1 cm^2 of the wall surface per second (fig. III.6). The volume of this parallelepiped is, obviously, equal to $v_x \text{ cm}^3$. It contains $dn v_x$ molecules whose velocity components lie between v_x and $v_x + dv_x$, v_y and $v_y + dv_y$, v_z and $v_z + dv_z$. The wall surface will be reached by all molecules contained in this parallelepiped, independently of the values of the velocity components v_y and v_z parallel to this surface.

The number of particles with given velocity component v_x (for arbitrary values of the two other components v_y and v_z) per unit volume is equal to

$$\begin{aligned} dn_{v_x} &= n(\alpha/\pi)^{\frac{3}{2}} e^{-\alpha v_x^2} dv_x \int_{-\infty}^{\infty} \int e^{-\alpha(v_y^2 + v_z^2)} dv_y dv_z = \\ &= n(\alpha/\pi)^{\frac{1}{2}} e^{-\alpha v_x^2} dv_x . \end{aligned}$$

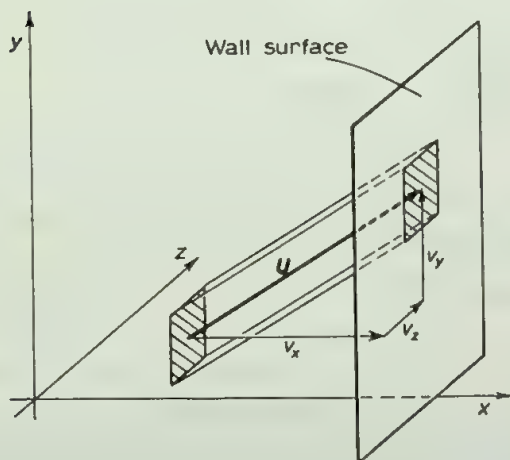


Fig. III.6

The number of particles contained in the parallelepiped of volume v_x is correspondingly equal to

$$dv = v_x \, dn_{v_x} = n(\alpha/\pi)^{\frac{1}{2}} e^{-\alpha v_x^2} v_x \, dv_x. \quad (8.1)$$

This expression gives the number of molecules having the given value of the velocity component v_x and reaching 1 cm² of the wall surface per second.

Each of dv molecules impinging on the wall transfers to it a momentum $2mv_x$, so that molecules with a given value of v_x transfer to the wall each second a momentum equal to $2mv_x dv$. Integrating this expression over all possible values of the velocity component v_x , we find the momentum transferred to 1 cm² of the wall surface per second by all the gas molecules striking it. This rate of transfer of momentum is, obviously, equal to the force acting on 1 cm² of the surface, i.e. to the gas pressure p

$$p = n(\alpha/\pi)^{\frac{1}{2}} \int_0^{\infty} 2mv_x e^{-\alpha v_x^2} v_x \, dv_x. \quad (8.2)$$

The integration in (8.2) is carried out only with respect to positive values of v_x , since molecules with negative values of the velocity component along the x -axis are moving away from the wall considered, not toward it*. It gives

$$p = \frac{m}{2\alpha} n, \quad (8.3)$$

or

$$p = \frac{m}{2\alpha} \frac{N}{V}. \quad (8.4)$$

To determine the numerical value of α it is necessary to compare eq. (8.4) with the experimental value of the pressure of a sufficiently rarefied gas. The latter is given by the equation of state

$$pV = NkT.$$

Comparing this expression with (8.4) we see that the parameter α is connected with the absolute temperature T by the relation

* In expression (8.2) collisions between molecules are not taken into account. However, molecules which do not reach the wall transfer their momentum to those reaching it.

$$\alpha = \frac{m}{2kT}. \quad (8.5)$$

The equality (8.5) confirms that the parameter α , which we introduced formally, is essentially a positive quantity. In determining the connection between α and the temperature T we had to resort to experimental data. In what follows we shall encounter a parameter analogous to α . The problem of the determination of the meaning and value of the parameter will then be discussed in more detail. We shall see that the meaning of α can be explained without directly involving experimental data.

Let us find the number ν of the impacts of molecules per cm^2 of the wall per second. Formula (8.1) gives the number of molecules reaching the wall per second and having a velocity between v_x and $v_x + dv_x$. The total number of molecules striking 1 cm^2 of the wall per second is obtained by integration of (8.1) over all values of v_x from zero to infinity. This gives

$$\nu = n(\alpha/\pi)^{\frac{1}{2}} \int_0^{\infty} e^{-\alpha v_x^2} v_x dv_x = n(kT/2\pi m)^{\frac{1}{2}}. \quad (8.6)$$

§9. Properties of the Maxwell distribution

Now we rewrite the Maxwell distributions (7.17) and (7.16), expressing the parameter α in terms of the absolute temperature of the gas

$$dn_v = n \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \exp \left[-\frac{m(v_x^2 + v_y^2 + v_z^2)}{2kT} \right] dv_x dv_y dv_z, \quad (9.1)$$

$$dn_v = 4\pi n \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \exp \left[-\frac{mv^2}{2kT} \right] v^2 dv. \quad (9.2)$$

Instead of making use of the state distribution function of gas particles, we can introduce the equivalent distribution, of the probabilities for a particle to be in a given state.

If the mean number of molecules having given velocity is equal to dn per cm^3 of the gas, while the total number of molecules is equal to n , then, obviously, the probability for an arbitrarily chosen molecule to be in a state with given velocity is equal to

$$dw = \frac{dn}{n}.$$

Hence the Maxwell distribution function can be treated as the function of distribution of the probabilities for a molecule to be in a given state. The latter is characterized by the values of the velocity components v_x , v_y and v_z .

Figs. III.7 and III.8 show, respectively, the functions of distribution of the probability density $n^{-1}(dn_{v_x}/dv_x)$ over the velocity components and $n^{-1}(dn_v/dv)$ over the absolute value of the velocity. Since it is impossible to show graphically a function of three variables, the quantity $n^{-1}(dn_{v_x}/dv_x)$ is plotted on the vertical axis, and the quantity $(m/2kT)^{1/2}v_x$ is plotted on the horizontal axis in fig. III.7.

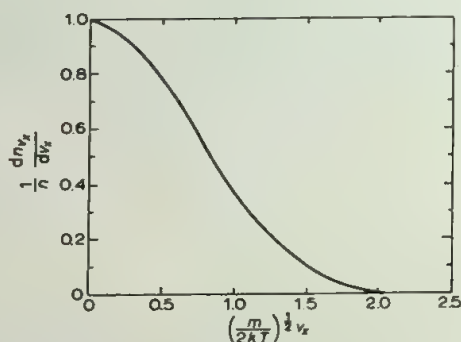


Fig. III.7

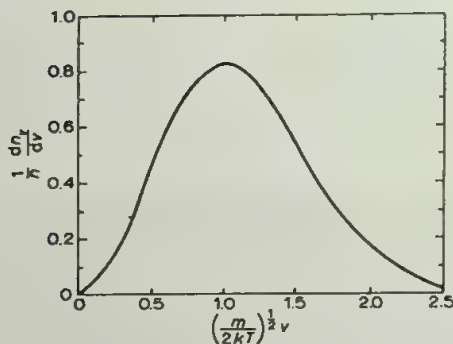


Fig. III.8

In fig. III.8 the function $n^{-1}(dn_v/dv)$ is plotted on the vertical axis, and the quantity $(m/2kT)^{1/2}v$ is plotted on the horizontal axis.

The numbers of molecules in a gas with very low and very high velocities turn out to be relatively small. Nevertheless, it is always possible to find a certain number of very fast and very slow molecules. At first sight it may seem strange that the velocity distribution function has a maximum. Indeed, the factor $\exp(-mv^2/2kT)$ decreases exponentially with the square of the velocity of the molecule. Hence the number of molecules with given velocity must be lower the higher the velocity. However, the second factor v^2 varies in the opposite direction and increases with increasing velocity. This factor characterizes the number of states of a molecule having a velocity lower than the given one. The competition of the two factors leads to the appearance of a maximum in the distribution function.

With increasing temperature the distribution becomes more and more gently sloping. This means that the relative number of molecules with a large value of the velocity increases progressively. Fig. III.9 shows the variation of the Maxwell distribution with increasing temperature (for oxygen molecules).

In addition to the velocity distribution one often has to make use of the momentum distribution and energy distribution of molecules.

Introducing into the distribution (7.17) the new variables

$$p_x = mv_x, \quad p_y = mv_y, \quad p_z = mv_z,$$

we find for the number of molecules with given momentum

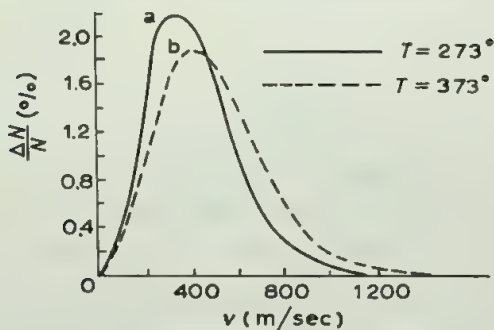


Fig. III.9

$$dn_p = n \left(\frac{1}{2\pi mkT} \right)^{\frac{3}{2}} \exp \left[-\frac{p_x^2 + p_y^2 + p_z^2}{2mkT} \right] dp_x dp_y dp_z. \quad (9.3)$$

Analogously, the number of molecules with absolute value of the momentum lying between p and $p+dp$ is equal to

$$dn_p = 4\pi n \left(\frac{1}{2\pi mkT} \right)^{\frac{3}{2}} \exp \left[-\frac{p^2}{2mkT} \right] p^2 dp. \quad (9.4)$$

Expressing the momentum in terms of the energy of the molecule $p = (2m\epsilon)^{\frac{1}{2}}$ we obtain for the number of particles with given energy

$$dn_\epsilon = n \frac{2}{[\pi(kT)^3]^{\frac{1}{2}}} e^{-\epsilon/\theta} \epsilon^{\frac{1}{2}} d\epsilon. \quad (9.5)$$

Experimental verification of the Maxwell distribution was one of the most important problems of molecular physics. Hence several methods of measuring the velocity distribution were devised. The most obvious of these is an experiment similar to the well-known Fizeau experiment on the determination of the velocity of light.

Molecules evaporating from the surface of a hot filament are let through a collimating system of slits, forming a narrow molecular beam travelling in vacuum in the direction of a cold trap. Two rotating discs with radial slits, through which the beam must pass, are placed in the path of the beam. In order that molecules having a velocity v_x may pass through both slits for a given angular velocity ω of the rotation of the discs and distance l between them, the slits must be displaced by an angle φ equal to

$$\varphi = \omega t = \omega \frac{l}{v_x}.$$

Thus, to each velocity v_x at given angular velocity ω there corresponds a definite displacement of the slits by an angle φ .

The number of particles in the beam having given velocity was determined directly from photometric measurement of the thickness of the deposit on the trap cooled by liquid air.

After evaluating the velocity distribution in the molecular beam corresponding to the velocity distribution in isotropic conditions, one obtains the histogram shown in fig. III.10. The curve in the drawing represents the Maxwell distribution (for mercury vapour).

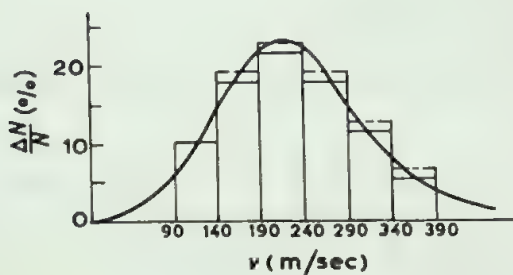


Fig. III.10

Another very accurate method of measuring the velocity distribution in the molecular beam of the vapour of lithium, sodium or similar gases is based on the investigation of the behaviour of the beam in a magnetic field perpendicular to the direction of motion of the beam. It represents a reproduction of the well-known Stern–Gerlach experiment for the determination of magnetic moments, but it cannot be described here in detail (see Part V).

§ 10. The calculation of characteristic quantities

Knowing the distributions (9.1) and (9.5), one can find the mean values of any quantities characterizing the properties of gas molecules. Let us find, first of all, the mean value of any velocity component, for example \bar{v}_x . By definition of the mean value

$$\bar{v}_x = \int v_x dw = \int v_x \frac{dn_v}{n} = \left[\frac{m}{2\pi kT} \right]^{\frac{3}{2}} \int_{-\infty}^{\infty} v_x \exp\left(-\frac{mv_x^2}{2kT}\right) dv_x \times \iint_{-\infty}^{\infty} \exp\left[-\frac{m(v_y^2 + v_z^2)}{2kT}\right] dv_y dv_z.$$

The integral with respect to dv_x , which is an odd function, is equal to zero. Hence the mean value of v_x is equal to zero:

$$\bar{v}_x = 0.$$

This result is quite obvious. It shows that both directions of motion along the x -axis are equally probable.

We would obtain a similar result in calculating any other velocity component.

We now find the mean value of the absolute value of the velocity. We have

$$\begin{aligned}\bar{v} &= \int v \frac{dn_v}{n} = 4\pi \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \int_0^{\infty} \exp \left[-\frac{mv^2}{2kT} \right] v^3 dv = \\ &= \frac{2}{\sqrt{\pi}} \left(\frac{2kT}{m} \right)^{\frac{1}{2}} = 1.13 \left(\frac{2kT}{m} \right)^{\frac{1}{2}}.\end{aligned}\quad (10.1)$$

In accordance with the foregoing the mean velocity of a molecule increases with increasing temperature. This increase is proportional to the square root of the absolute temperature of the gas. We see also that the mean velocity of a molecule is inversely proportional to the square root of the mass of the molecule.

Of great interest is the mean value $\bar{\epsilon}$ of the kinetic energy of a gas molecule. From (9.5) it is equal to

$$\bar{\epsilon} = \int \epsilon \frac{dn_{\epsilon}}{n} = \frac{2}{\sqrt{\pi}(kT)^{\frac{3}{2}}} \int_0^{\infty} \exp [-\epsilon/kT] \epsilon^{\frac{3}{2}} d\epsilon.$$

In calculating the above integral it is necessary to introduce a new variable $x = \epsilon^{\frac{1}{2}}$. A simple calculation leads to the formula for the mean value of the kinetic energy of the translational motion of a molecule.

$$\bar{\epsilon} = \frac{3}{2} kT. \quad (10.2)$$

It is seen that the mean energy of a molecule does not depend on its nature and is proportional to the gas temperature T . The mean energy E of all gas molecules in a container is equal to the sum of the energies of the translational motion of all the molecules, since there is no interaction between them:

$$E = N\bar{\epsilon} = \frac{3}{2} NkT. \quad (10.3)$$

where N is the total number of molecules in the gas.

The energy of a given portion of an ideal gas does not depend on the volume of the container and is determined only by the absolute temperature *.

We shall identify this mean energy of the mechanical motion of gas molecules with the macroscopic thermal energy. In this connection we must treat the absolute temperature, from the kinetic point of view, as a quantity characterizing the mean energy of the motion of molecules. At present such a treatment is the only possible one. In the following chapter we shall discuss these statements in detail.

The expressions obtained for the mean energy of a molecule and for the gas as a whole can be interpreted in the following way. Every molecule has three translational degrees of freedom, and its motion can be resolved into the motion in three mutually perpendicular directions. By virtue of the equivalence of all directions in space, the mean energy of motion in each direction must be the same. Thus, formula (10.2) means that there corresponds to each degree of freedom of the motion, on the average, an energy equal to $\frac{1}{2}kT$. This statement is a particular case of the general law of equipartition of energy over degrees of freedom. In §39 we shall discuss this law in detail and point out the limits of its applicability.

It is of interest to establish the connection between the energy ϵ of the gas and its pressure p .

Writing the expression (8.2) in the form

* It may seem that the independence of the mean energy of a gas molecule on the size of the container contradicts the quantum formula (1.11) for the energy, according to which $\epsilon_n \sim 1/a^2$, where a is the linear dimension of the container. However, it should be borne in mind that the mean energy is determined by the integral of the product of the energy and the number of quantum states. The latter is, according (1.24), proportional to $\Omega(\epsilon)\Delta\epsilon \sim V\sqrt{\epsilon}\Delta\epsilon \sim a^3(\Delta\epsilon/a) \sim a^2\Delta\epsilon$. This leads to the independence of the energy on the volume.

$$\begin{aligned}
 p &= 2n \left(\frac{m}{2\pi kT} \right)^{\frac{1}{2}} \int_0^{\infty} \exp \left[-\frac{mv_x^2}{2kT} \right] (mv_x^2) dv_x = \\
 &\quad \frac{4n}{3} \left[\left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \int_0^{\infty} \frac{mv_x^2}{2} \exp \left[-\frac{mv_x^2}{2kT} \right] dv_x \int_{-\infty}^{\infty} \exp \left[-\frac{m(v_y^2+v_z^2)}{2kT} \right] dv_y dv_z \right. \\
 &\quad + \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \int_0^{\infty} \frac{mv_y^2}{2} \exp \left[-\frac{mv_y^2}{2kT} \right] dv_y \int_{-\infty}^{\infty} \exp \left[-\frac{m(v_x^2+v_z^2)}{2kT} \right] dv_x dv_z \\
 &\quad \left. + \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \int_0^{\infty} \frac{mv_z^2}{2} \exp \left[-\frac{mv_z^2}{2kT} \right] dv_z \int_{-\infty}^{\infty} \exp \left[-\frac{m(v_x^2+v_y^2)}{2kT} \right] dv_x dv_y \right] = \\
 &= \frac{2n}{3} \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \int_{-\infty}^{\infty} \frac{mv^2}{2} \exp \left[-\frac{mv^2}{2kT} \right] dv_x dv_y dv_z = \\
 &= \frac{2}{3} \frac{N}{V} \left(\frac{mv^2}{2} \right) = \frac{2}{3} \frac{N\bar{\epsilon}}{V} = \frac{2}{3} \frac{E}{V}, \tag{10.4}
 \end{aligned}$$

we see that the pressure of an ideal gas turns out to be numerically equal to $\frac{2}{3}$ of the kinetic energy of the translational motion of gas molecules contained in unit volume.

Finally, the velocity of gas molecules at which the Maxwell distribution has a maximum, i.e. the most probable velocity $v_{m.p.}$, is determined from the condition

$$\frac{d}{dv} \left(\exp \left[-\frac{mv^2}{2kT} \right] v^2 \right) = 0.$$

We easily find

$$v_{m.p.} = \left(\frac{2kT}{m} \right)^{\frac{1}{2}}. \tag{10.5}$$

Comparing (10.5) with (10.1) we see that the mean velocity of the molecules is 13% higher than the most probable velocity.

The notion of the root-mean-square velocity $(\overline{v^2})^{\frac{1}{2}}$, which characterizes the energy of gas molecules, is also often introduced. By virtue of (10.2) this quantity is equal to

$$(\overline{v^2})^{\frac{1}{2}} = \left(\frac{3}{2}\right)^{\frac{1}{2}} \left(\frac{2kT}{m}\right)^{\frac{1}{2}} = 1.22 \left(\frac{2kT}{m}\right)^{\frac{1}{2}}. \quad (10.6)$$

The root-mean-square velocity is 22% higher than the most probable velocity. This is quite natural, since the contribution of fast molecules to the energy must be larger than that of slow molecules.

§11. Collisions of molecules with each other

Consider two molecules moving in an ideal gas with the velocities \mathbf{v}_1 and \mathbf{v}_2 . It is obvious that for the collision of these molecules with each other the absolute values and directions of the velocities do not in themselves play any role. The relative motion of one molecule with respect to the other only is of importance. If, for example, both molecules are moving on a straight line one following the other, then a collision will take place in unit time if the second molecule manages to "reach" the first molecule in 1 second. The velocities of motion of the two molecules in space with respect to the walls of the container are not relevant.

Thus, in solving the problem of collisions it is necessary to consider their relative motion. In §42 of Part I it has been shown that the motion of two particles can always be resolved into the motion in space of the common centre of mass and their relative motion.

We write the probability that the first molecule has a velocity \mathbf{v}_1 and the second \mathbf{v}_2 in the form

$$\begin{aligned} dw_{12} &= dw_1 dw_2 = \\ &= \left(\frac{m_1}{2\pi kT}\right)^{\frac{3}{2}} \exp\left[-\frac{m_1 v_1^2}{2kT}\right] d\mathbf{v}_1 \left(\frac{m_2}{2\pi kT}\right)^{\frac{3}{2}} \exp\left[-\frac{m_2 v_2^2}{2kT}\right] d\mathbf{v}_2. \end{aligned}$$

We pass over from the variables \mathbf{v}_1 and \mathbf{v}_2 to new variables \mathbf{R} , \mathbf{v}_{rel} , making use of formulae (42.4) of Part I. Then we have

$$m_1 v_1^2 + m_2 v_2^2 = M \dot{R}^2 + \mu v_{\text{rel}}^2, \quad d\mathbf{v}_1 d\mathbf{v}_2 = |I| d\dot{\mathbf{R}} dv_{\text{rel}},$$

where $\mu = m_1 m_2 / (m_1 + m_2)$ is the reduced mass, $M = m_1 + m_2$, and $|I|$ stands for the determinant of the Jacobian of the transformation from $\mathbf{v}_1, \mathbf{v}_2$ to $\dot{\mathbf{R}}, \mathbf{v}_{\text{rel}}$

$$|I| = \begin{vmatrix} \frac{\partial \mathbf{v}_1}{\partial \mathbf{v}_{\text{rel}}} & \frac{\partial \mathbf{v}_1}{\partial \dot{\mathbf{R}}} \\ \frac{\partial \mathbf{v}_2}{\partial \mathbf{v}_{\text{rel}}} & \frac{\partial \mathbf{v}_2}{\partial \dot{\mathbf{R}}} \end{vmatrix} = \begin{vmatrix} \frac{m_2}{m_1 + m_2} & 1 \\ -\frac{m_1}{m_1 + m_2} & 1 \end{vmatrix} = 1.$$

Thus,

$$\begin{aligned} dw_{12} &= \left(\frac{m_1 m_2}{2\pi(m_1 + m_2)kT} \right)^{\frac{3}{2}} \exp \left[-\frac{\mu v_{\text{rel}}^2}{2kT} \right] d\mathbf{v}_{\text{rel}} \left(\frac{m_1 + m_2}{2\pi kT} \right)^{\frac{3}{2}} \exp \left[-\frac{M \dot{R}^2}{2kT} \right] d\dot{\mathbf{R}} = \\ &= dw_{\text{rel}} dw_{\text{c.m.}} \end{aligned}$$

We see that the probability of a given state of motion of two particles is equal to the product of the probabilities of two independent events: the probability for the particles to have a given relative velocity,

$$dw_{\text{rel}} = \left(\frac{\mu}{2\pi kT} \right)^{\frac{3}{2}} \exp \left[-\frac{\mu v_{\text{rel}}^2}{2kT} \right] dv_{\text{rel}}, \quad (11.1)$$

and the probability for the centre of mass of the system of the two particles to move in space with a given velocity

$$dw_{\text{c.m.}} = \left(\frac{M}{2\pi kT} \right)^{\frac{3}{2}} \exp \left[-\frac{M \dot{R}^2}{2kT} \right] d\dot{\mathbf{R}}.$$

In the problem of collisions, only the first probability is of interest. In the case of particles with the same mass $\mu = \frac{1}{2}m$, so that

$$dw_{\text{rel}} = \left(\frac{m}{4\pi kT} \right)^{\frac{3}{2}} \exp \left[-\frac{mv_{\text{rel}}^2}{4kT} \right] dv_{\text{rel}}. \quad (11.2)$$

By means of the probability distribution (11.2) the mean value \bar{v}_{rel} of the velocity of relative motion can be found

$$\bar{v}_{\text{rel}} = 4\pi \left(\frac{m}{4\pi kT} \right)^{\frac{3}{2}} \int_0^{\infty} \exp \left[-\frac{mv_{\text{rel}}^2}{4kT} \right] v_{\text{rel}}^3 dv_{\text{rel}} = \sqrt{2} \bar{v}. \quad (11.3)$$

Thus, the mean velocity of relative motion is almost one and a half times the mean velocity of the thermal motion.

Let us now find the number of collisions undergone per unit time by a molecule of a gas having a number density n . We assume that the gas is so rarefied that the molecules only collide in pairs, that is, collisions in which three or more molecules come simultaneously into direct contact can be disregarded. The process of collision of molecules can be characterized by their cross section σ .

In considering collisions between gas molecules we shall assume that all the gas molecules except one are at rest. The molecule singled out moves relative to those at rest with a velocity v_{rel} . In unit time it traverses a path v_{rel} and collides with all particles lying in a cylinder of volume σv_{rel} . The number of such collisions is, obviously, equal to $\sigma v_{\text{rel}} dn_{v_{\text{rel}}}$, where $dn_{v_{\text{rel}}} = n dw_{\text{rel}}$ and dw_{rel} is given by formula (11.2).

The total number of collisions undergone by the molecule per unit time is obtained by integrating this expression over all possible values of v_{rel} :

$$\begin{aligned} \nu &= \int \sigma v_{\text{rel}} dn_{v_{\text{rel}}} = \\ &= 4\pi \left(\frac{m}{4\pi kT} \right)^{\frac{3}{2}} n \int_0^{\infty} \sigma(v_{\text{rel}}) \exp \left[-\frac{mv_{\text{rel}}^2}{4kT} \right] v_{\text{rel}}^3 dv_{\text{rel}}. \end{aligned} \quad (11.4)$$

If the cross section for the collision can be assumed to be independent of the velocity, then instead of (11.4) we obtain

$$\begin{aligned}
 \nu &= 4\pi \left(\frac{m}{4\pi kT} \right)^{\frac{3}{2}} n\sigma \int_0^{\infty} \exp \left[-\frac{mv_{\text{rel}}^2}{4kT} \right] v_{\text{rel}}^3 dv_{\text{rel}} = \\
 &= n\sigma \bar{v}_{\text{rel}} = n\sigma \sqrt{2} \bar{v} = 4n\sigma \left(\frac{kT}{\pi m} \right)^{\frac{1}{2}}.
 \end{aligned} \tag{11.5}$$

This is just the number of collisions undergone by the molecule per second.

§ 12. The mean free path

Let us now find the mean path traversed by a molecule between two successive collisions, which is called the mean free path.

In one second a molecule traverses in space a path equal, on the average, to \bar{v} , and undergoes ν collisions. The mean path between collisions is equal to

$$\lambda = \frac{\bar{v}}{\nu} = \frac{\bar{v}}{n\sigma \bar{v}_{\text{rel}}} = \frac{1}{n\sigma \sqrt{2}}. \tag{12.1}$$

The path λ is on the average proportional to the ratio of the mean path traversed per unit time by the molecule to the number of collisions undergone by it. Hence the quantity λ is called the mean free path. The mean free path λ turns out to be inversely proportional to the gas density n and the cross section σ .

Formula (12.1) gives the mean free path. However, it is often important to know the probability for a molecule to traverse an arbitrary path x without undergoing any collision. In other words, what is of interest here is the law of the probability distribution of free paths for the molecules. We denote by $w(x)$ the probability for a molecule to traverse a distance x without undergoing any collision.

Correspondingly, $w(x+dx)$ represents the probability for a molecule to traverse a path $x+dx$ without undergoing any collision.

The traversal of the path $x+dx$ is a complex event, consisting of two independent stages: the traversal of the path x without collisions and the subsequent traversal of the path dx also without collisions.

Since these events are independent, we can write that

$$w(x+dx) = w(x) w(dx). \tag{12.2}$$

It is convenient to rewrite the above probability in another form. It is obvious that the probability $w'(dx)$ for a molecule to undergo a collision on the infinitesimally small path dx is proportional to the length dx and can be written in the form adx , where a is a certain coefficient of proportionality. The probability that a molecule will traverse the path dx without collisions is equal to

$$w(dx) = 1 - w'(dx) = 1 - adx .$$

Substituting this into (12.2) we find

$$w(x+dx) = w(x) (1 - adx) . \quad (12.3)$$

Expanding $w(x+dx)$ in a series in powers of dx and confining ourselves to infinitely small quantities of the first order of smallness, we have

$$w(x+dx) = w(x) + \frac{dw}{dx} dx ,$$

whence, substituting into (12.3), we obtain

$$dw = -aw(x) dx .$$

Integrating, we find

$$w(x) = Ae^{-ax} .$$

In order to determine the arbitrary constant A , we note that the probability for a molecule to traverse an arbitrarily small path without collisions is equal to unity:

$$w(x \rightarrow 0) = 1 .$$

Whence it follows that $A = 1$ and, finally,

$$w(x) = e^{-ax} . \quad (12.4)$$

In order to determine the meaning of the constant a we shall find the mean free path λ , making use of formula (12.4).

By definition the mean free path λ is equal to

$$\lambda = \int_0^{\infty} x dP, \quad (12.5)$$

where dP is the probability that a molecule, after traversing a path x , without collisions, will undergo a collision on the segment $x, x+dx$. According to the foregoing, we can write

$$dP = w(x) w'(dx) = w(x) a dx = a e^{-ax} dx. \quad (12.6)$$

Substituting the value of dP into (12.5) we find

$$\lambda = a \int_0^{\infty} x e^{-ax} dx = a^{-1}. \quad (12.7)$$

Thus, the constant a turns out to be a quantity inversely proportional to the mean free path.

Formula (12.4) for the probability that a molecule will traverse a path x without undergoing any collision can be rewritten in the form

$$w(x) = e^{-x/\lambda}. \quad (12.8)$$

This probability turns out to be an exponentially decreasing function of the distance. It should be stressed that $w(x)$ gives the probability for a molecule to traverse a path x without collisions irrespective of the position where it underwent the last collision. This means that the distance x is measured from an arbitrary point and not from the position of the last collision.

The probability for a molecule to traverse a path without collisions and then to collide in the segment $x, x+dx$ is, according to (12.6) and (12.7), equal to

$$dP = \lambda^{-1} e^{-x/\lambda} dx. \quad (12.9)$$

Formula (12.8) is important for the experimental determination of the mean free path of gas molecules. Imagine a narrow collimated molecular beam entering a container, evacuated to a relatively low pressure, containing a cooled plate placed in the path of the beam, first at a distance x_1 and then at a distance x_2 from the inlet. Molecules which traverse the paths x_1 and x_2

without collisions will reach the plate, forming a deposit on it. The ratio of the number of particles deposited on the plate in the two positions is, according to (12.8), equal to

$$\frac{N(x_1)}{N(x_2)} = \frac{e^{-x_1/\lambda}}{e^{-x_2/\lambda}}. \quad (12.10)$$

By measuring the numbers $N(x_1)$ and $N(x_2)$ and assuming that λ is the same for all molecules of the beam, λ can be determined by means of (12.10).

In accordance with the requirements of the theory, λ turns out to be inversely proportional to the density or, what is the same, to the pressure of the gas. The order of magnitude of λ amounts to about 10 cm at $p \approx 10^{-3}$ mm of mercury and about 10^{-5} cm at atmospheric pressure.

Statistical Distribution

§ 13. Quasi-independent systems

Consider a macroscopic system consisting of a very large number of particles. We assume that the motion of all the particles is determined by the laws of quantum mechanics; the state of each particle is characterized by certain quantum numbers. We divide the whole system into a large number of parts in such a way that the interaction between these parts is very weak and can be neglected in the first approximation. We shall call the whole system under investigation the ensemble of its almost independent parts.

These weakly interacting parts of the system are, in the first approximation, moving independently of one another. However, the interaction existing between them leads to the fact that, in reality, the motion of one of the parts affects the motion of another, and that they are not completely independent. In the following we shall call the weakly interacting parts of a large system the quasi-independent subsystems, or simply subsystems.

We first of all dwell on the problem for which the subsystems forming a large system can be considered as independent.

It is obvious that the subsystems are quasi-independent if the energy of their interaction is, on the average, small in comparison with the energy of each of the subsystems. This means that in some cases the interaction between

the subsystems can be relatively large, but the duration of the interaction must be so small that they spend an overwhelmingly large part of the time without interacting with each other.

As an example of such subsystems one can mention the molecules of an ideal gas, which interact strongly with one another only seldom and then for a very short time interval.

In other cases a continuous, but weak, interaction can exist between subsystems. Imagine, for example, that each of the subsystems of the system contains a very large number of particles (atoms or molecules) and, thus, represents a macroscopic system. Then the total energy of a subsystem, made up of the energies of motion of the individual particles, is proportional to the total number of particles of the subsystem. The number of particles is in its turn proportional to the volume of that subsystem. The interaction between various subsystems is mainly due to the forces of molecular interaction between the molecules which are at the surface of each of the interacting subsystems *. These forces decrease so rapidly with distance that the contribution to the energy of interaction of molecules which are deep inside the subsystems is small in comparison with that of the surface molecules. Hence the energy of interaction between the subsystems is proportional to the number of molecules at their surface, i.e. to the size of the surface.

Thus, the energy $\bar{\epsilon}$ of a subsystem is proportional to R^3 , where R is the characteristic linear dimension of the system, and the interaction energy is $\bar{\epsilon}_{\text{int}} \sim R^2$. Their ratio

$$\frac{\bar{\epsilon}_{\text{int}}}{\bar{\epsilon}} \sim \frac{R^2}{R^3} \sim \frac{1}{R} \sim N^{-\frac{1}{3}}$$

becomes negligibly small for a sufficiently large N .

The energy of the entire ensemble of quasi-independent systems can then be assumed to be equal to the sum of the energies of the individual parts, i.e.

$$E \approx \sum \epsilon_i, \quad (13.1)$$

* We shall nowhere take into account the gravitational interaction between the molecules, because it is very weak and does not play any part in molecular processes. However, it should be noted that, in studying the macroscopic properties of matter in astrophysical problems, the gravitational field is in a number of cases of essential importance and must be taken into account.

where the sign \approx underlines the fact that in writing (13.1) we have disregarded the energy of interaction between the subsystems forming the ensemble. The summation in (13.1) is carried out over all parts of the system (the subsystems).

§ 14. Statistical distribution

From the whole system we single out mentally an arbitrarily chosen subsystem. This subsystem consists of a certain number of molecules (as we have just explained, it can be either small or large, depending on the actual nature of the subsystems constituting the system) moving according to the laws of quantum mechanics. The energy of our subsystem is not strictly constant, but, on the contrary, varies continuously within the range of the value ϵ_{int} , where ϵ_{int} is the energy of the interaction of the system with its surroundings. Although ϵ_{int} is very small and can be disregarded in the energy balance, it nevertheless plays an essential part in the behaviour of the system. The interaction of the subsystem with bodies surrounding it is the cause of its transitions from one quantum state into others. This interaction is of an extremely complicated character. Even in the simplest case, when individual molecules were taken as subsystems, we saw that an attempt to determine the motion of each molecule, i.e. the sequence of change of states, presented great difficulties. In a gas consisting of a very large number of particles new laws are introduced which we have briefly formulated in the proposition of molecular disorder. The situation is completely analogous also in the general case of the macroscopic system consisting of a large number of quasi-independent subsystems. The interaction between the subsystems is so complex that an accurate determination of the state of each of the subsystems becomes a problem even more difficult than finding the motion of individual gas molecules. At the same time, such a determination loses any physical meaning. Indeed, even if we succeeded in the determination of the state in which a certain subsystem is at a given instant, it would, as a result of the interaction with other subsystems, pass over into another state in a very short time. Hence the particular known state of a subsystem entering into a large ensemble of systems, at a certain instant, does not characterize the overall ensemble, just as the velocity of an individual gas molecule does not characterize the properties of the gas as a whole. In this connection we renounce beforehand any attempt to describe the behaviour of an individual subsystem, and instead look for the statistical laws characterizing the behaviour of the overall ensemble of the subsystems as a whole. This means that we shall not try to

trace in detail the successive changes of state of an individual subsystem in the course of time, but shall seek to find the probability w_i for an arbitrarily chosen subsystem of the ensemble to get into a certain i th state. If we find this distribution, we shall then be able:

(1) *to find the mean number of subsystems* in a given state, if an ensemble consisting of N identical subsystems is given (for example, the number of molecules in a given state, if the subsystems correspond to individual gas molecules);

(2) *to find the mean value of any quantity characterizing the state of an individual system*, for example its energy, from the general rules set out in §4;

(3) *to find the deviations of quantities from their mean values*, characterized by the mean square deviation.

In this case we can apply the general statistical considerations of §5 to a macroscopic quasi-independent subsystem consisting in its turn of a very large number of particles. These considerations show that in such a system all quantities have values differing very little from their mean values. Hence if we know the latter, then we can assume, with a high degree of accuracy, that we know the true values of all the quantities characterizing a state of a subsystem.

Thus we see that the statement of the problem in statistical physics differs in no way in principle from that in the kinetic theory of gases. We shall investigate the statistical laws shown in systems consisting of a very large number of particles. Knowing these laws, we can calculate the mean values of all sorts of quantities. Since, however, the objects in which we are interested are macroscopic bodies consisting of a very large number of particles, the predictions obtained from statistical laws have a completely reliable character. The mean values of all the quantities are the same as the true values to a high degree of accuracy.

However, besides this similarity between the general statement of the problem in statistical physics and in the kinetic theory of gases there is also a fundamental difference between them. In the kinetic theory of gases an individual quasi-independent system was always a molecule of a rarefied gas. The molecule was assumed to be monatomic, since only its translational motion was considered. The gas as a whole corresponded to our ensemble of systems. In statistical physics the problem is stated much more broadly. An individual subsystem can be any quasi-independent system. It can be a monatomic molecule in a rarefied gas as well as a polyatomic molecule performing not only a translational motion but also a rotational and vibratory motion. The system can also be a gas as a whole, confined to a certain con-

tainer. The walls of this container and the bodies surrounding it play the role of other systems with which the gas in the container (a quasi-closed system) interacts weakly and exchanges energy. The gas, the walls of the container and the bodies surrounding it form an ensemble of subsystems. A subsystem can, for example, be a solid body containing a sufficiently large number of particles. The bodies surrounding it play the role of other parts of the ensemble.

Thus, an ideal gas, considered in the kinetic theory of gases, is a particular and the simplest case of a general statistical system.

In the preceding chapter we have, for the particular case of an ideal gas, partly fulfilled the programme which we outlined at the beginning of this section. We saw that the stationary probability density distribution of different states of a molecule in the gas is established because of the interaction between gas molecules in collisions.

In the same way, in the general case of an ensemble of arbitrary quasi-stationary systems, a certain distribution of the probabilities for a subsystem to get into a definite energy state ϵ_i will be established. In the following sections the derivation of the statistical probability distribution for an arbitrary subsystem will be developed.

§ 15. The probability of a state of a system

Let us observe mentally the changes of state of a subsystem which we have singled out arbitrarily. We shall, for brevity, call all other parts of the system, constituting the surroundings of this subsystem, the reservoir. The meaning of this term will be clear from what follows. We shall call the subsystem itself, if not specified otherwise, simply the system. Each state of the system is characterized by a set of quantum numbers. If the system has f degrees of freedom, then its state is characterized by a set of f quantum numbers.

To each set of quantum numbers there corresponds a certain quite definite energy of the system*. If the system consists of a large number of particles and has many degrees of freedom, different energy levels corresponding to different, but mutually similar sets of quantum numbers, lie very close to one another. In the limiting case where the number of particles is very large,

* In what follows we shall dwell in more detail on this problem and take into account the case where different sets of values of the quantum numbers correspond to one and the same energy, i.e. the case of degenerate systems.

so that the system is macroscopic, we pass over from a quantum system to a classical system *. In this case all the energy levels merge into a continuous spectrum, and instead of discrete levels use can be made of the continuous energy range of the classical theory.

As we have already stressed, because of the interaction with the surroundings the energy of the system is not constant. Hence it is senseless to speak about a well defined energy of the system, and it should be said only that its energy lies within ϵ and $\epsilon + \delta\epsilon$. To energy values of the system lying within the range ϵ and $\epsilon + \delta\epsilon$ there corresponds a certain number $\Omega(\epsilon)\delta\epsilon$ of quantum states. We shall often call $\Omega(\epsilon)$ the number of quantum states corresponding to the energy ϵ of the system, or the degeneracy of a given state. This should not lead to misunderstanding. In fact, we shall assume the above definition. It is obvious that to different energy values ϵ there corresponds a different number $\Omega(\epsilon)$ of quantum states. It is different also for different physical systems.

In the case where the system is an individual atom or molecule, the number of quantum states corresponding to a given energy is small for small excitation energies, but increases rapidly with increasing energy. If the system is a macroscopic body, there is always a practically continuous energy spectrum.

In what follows we shall make use of the actual value of $\Omega(\epsilon)$ for the simplest systems, as derived in ch. 1.

As a result of complex and random interactions between the system and its surroundings (the reservoir) the states of the system will change and it will go over from one quantum state into others. The system will perform transitions between different states corresponding to a given energy value ϵ (more precisely, corresponding to an energy lying within the range ϵ , $\epsilon + \delta\epsilon$) as well as between states with different energies ϵ_1 , ϵ_2 ,

For example, in the case where the subsystem is an individual molecule, collisions with other molecules and the wall of the container, forming the thermostat, lead to transitions into other states with the same energy (changes in the direction of motion only) or into states with a different energy (inelastic collisions or elastic collisions with a relatively large momentum transfer).

If one observes the change of the system during a relatively long time interval, then it will pass through all possible states. The state of the system at a given instant will depend neither on its initial state nor on the initial state of the reservoir. The effect of initial conditions will be completely concealed

* The problem of the transition to classical systems will be considered more precisely in Part V.

by complicated transitions, interactions, and so on. Hence we can say that the state of the system at any instant will be determined by the complex pattern of random interaction between the system and its surroundings.

If attention is first drawn to transitions between different states belonging to a given energy value (between ϵ and $\epsilon + \delta\epsilon$), then it is physically obvious that all these states are mutually equivalent and that no one can have any advantages over others. The equivalence of the states of the system belonging to a given energy is a generalization of the proposition of molecular disorder in an ideal gas. In fact, the latter means that all states with the same energy but with different directions of motion in space and different positions in the container are equally probable in an ideal gas.

If the system is observed during a sufficiently long time interval, then, since all states with a given energy are equivalent, it will appear in all these states, irrespective of the state which it was in initially. Moreover, since the transition of the system from one state into another occurs as a result of accidental perturbations and the action of its surroundings and all quantum states belonging to a given energy are completely equivalent, it can be said that the system will have an equal chance to get into each of them. Hence the time during which the system is in each of the quantum states belonging to a given energy is the same for all these states. To characterize a state, one usually gives not the time interval during which the system is in it, but the ratio of this time to the overall observation time, i.e. the probability of realization of a given state. Then the preceding statement can be formulated briefly as the following principle: in a quasi-closed system all quantum states belonging to a given energy (lying between ϵ and $\epsilon + \delta\epsilon$) are equivalent. This statement is called the law of equal probability of elementary quantum states.

The question as to whether, in fact, a macroscopic system can be in all states with a given energy, without exceptions, has been the object of discussion of physicists and mathematicians over the course of a number of years.

Systems which can enter any state of a given energy during a sufficiently long time interval are called ergodic systems. If at the initial instant the ergodic system is in a certain state, then sooner or later it will get into any other state previously chosen out of a given group of states (ergodic hypothesis). Although this proposition seems to be very likely, there is no proof for it. Difficulties arising in connection with this hypothesis are overcome in quantum statistics by considering idealized systems in which transitions into some of the states are forbidden by selection rules. These rules are understood to be certain restrictions of the possibility of transitions due to various

causes, which we cannot go now into *. Then all the microscopic states can be divided into two groups — those between which transitions are possible, and those into which systems of the first group cannot get. If the second group of states is indeed completely forbidden, then, in considering the properties of the system, the states of this group can be assumed to be non-existent. On the contrary, the system sooner or later gets into all states of the first group.

In reality, there are no such idealized systems in nature. Also, more or less probable transitions between any states of the system are possible. The probabilities of these transitions can be considerably different from each other. Imagine, for example, a system consisting of atoms which can be in different states and which can interact with one another. Various processes can take place in the system as a result of this interaction, provided they do not contradict the basic laws of motion. For example, a fraction of the atoms can obtain energy from other atoms and pass over into an ionized state. However, such processes occur with such a low probability and, hence, proceed so slowly that in considering the behaviour of the system during any practically feasible time interval in terrestrial conditions, we can completely disregard the possibility of their occurrence.

The same thought can be expressed as follows: the transition of the system into a state corresponding to bare nuclei and free electrons is forbidden, and only the states of the system consisting of atoms should be considered. Since our purpose is the calculation of mean values, and the contribution of highly improbable states to them is very small, we shall not commit any significant error by ignoring completely transitions into forbidden states which are possible, but improbable. If it is assumed that transitions are possible between other (allowed) states and that the system spends a sufficiently long time in all these states (in this restricted form the ergodic hypothesis is physically sufficiently convincing), then the principle of equal probability of all microscopic states can be illustrated by the following considerations.

Assume that there is a system of particles which can be in states 1, 2, ..., i , ... with one and the same energy. Let us denote by N_i the number of particles in the i th state and by N_k the number of particles in the k th state. Further, let w_{ik} be the probability of transition of a particle from the i th state into the k th state. This probability can be calculated by means of the laws of quantum mechanics. Finally, let w_{ki} be the probability of transition from the k th state into the i th state.

In quantum mechanics it turns out that all processes taking place with

* See Part V, § 106.

individual microscopic particles are strictly reversible, so that the probabilities of a direct transition w_{ik} and a reverse transition w_{ki} are always equal to each other*. This is the so-called principle of microscopic reversibility. The number of particles passing over from the i th state into the k th state per unit time is, obviously, equal to the number N_i of particles in the i th state multiplied by the probability w_{ik} of the transition. The number of reverse transitions is correspondingly equal to $N_k w_{ki}$. In the stationary state the number of direct and reverse transitions must be the same, since the system must remain on the average in a steady state.

$$N_i w_{ik} = N_k w_{ki}, \quad (15.1)$$

but, by virtue of the principle of microscopic reversibility $w_{ik} = w_{ki}$, whence it follows that $N_i = N_k$.

This reasoning can be extended to all other states between which there are transitions (i.e. $w_{ik} \neq 0$). It turns out that the numbers of particles in all states must be equal to one another and, consequently, all quantum states must be equally probable.

If we now pass over to transitions between states of the system with different energy (more precisely, with an energy differing by an amount much larger than the interaction energy ϵ_{int}), then it can be stated that, because of the open character of the system, it can also make such transitions.

In transitions into states with a higher energy the energy difference will be drawn by the system from its surroundings. In transitions into states with a lower energy the energy excess will be transferred to its surroundings. The probability of such transitions is ensured by the existence of an interaction. It is just the interaction that is the cause of transitions of the system from one state into others. These considerations are a generalization of the principle of molecular disorder, which we have used in the kinetic theory of gases. However, it should be stressed that the reasoning given is more physically likely than strictly substantiated. Hence the ergodic hypothesis should be assumed as a postulate the validity of which is proved by comparison of theory with experiment. In any case, for statistics it is sufficient that this hypothesis be fulfilled approximately in most cases.

* See Part V, §56.

§ 16. The Gibbs distribution

We now ask the question: what is the probability w_i of finding our system in states with an energy lying between ϵ_i and $\epsilon_i + \delta\epsilon_i$ (where $\delta\epsilon_i \ll \epsilon_i$, and the subscript i runs through the values 1, 2, 3, ...)? To each value of the energy ϵ_i there corresponds a certain group $\Omega(\epsilon_i)$ of quantum states.

Consider at first a closed system which does not interact with the bodies surrounding it. In reality, no perfectly closed systems can exist in nature. Whatever the physical nature of the system may be, it always interacts, if only very weakly, with the bodies surrounding it. In quantum mechanics it turns out that a system can have a strictly constant energy only when it is in the ground state (for a macroscopic system this corresponds to the state at absolute zero; see § 34 of Part V).

Hence we shall understand a closed system to be a system whose energy during the entire time of observation remains within a given narrow range $\delta\epsilon_i$.

Since all states with a given energy are equally probable, the probability that a closed system be found in one of the states with given energy is simply proportional to the number of states with that energy:

$$w(\epsilon_i) \sim \Omega(\epsilon_i). \quad (16.1)$$

Formula (16.1) is called the Gibbs microcanonical distribution. The microcanonical distribution shows that the probability for a closed system to be in a state with given energy is proportional to its degeneracy.

The Gibbs microcanonical distribution is the basis of statistical physics. It shows that a closed system will be found with greater probability in a state with a higher degeneracy.

In phase space the states of a closed system lying in a narrow interval $\delta\epsilon$ form a very thin layer, which reduces to a constant-energy surface as $\delta\epsilon \rightarrow 0$. Each cell in the constant-energy layer (surface) corresponds to each possible quantum state.

As an example of the application of the microcanonical distribution we shall consider a system of N particles which do not interact with one another and which can be in two different states. Specifically, we shall speak of particles with a spin one half (in units of $\hbar/2\pi$). In this case the projection of the spin of every one of the particles onto an arbitrary axis can take on two values: $s_{\uparrow} = \frac{1}{2}$ and $s_{\downarrow} = -\frac{1}{2}$. We shall conditionally call these spins upward and downward directed respectively. In the absence of an external magnetic field the energy of the system depends neither on the orientation of the spins of

the particles nor on that of the total spin $S = \sum s$ of the system. Hence to a given value of the energy of the system ϵ there corresponds a large number of different states corresponding to different orientations of the spins of the individual particles.

According to what has been said, all states with a given distribution of spin orientation are equally probable.

We apply formula (16.1) for finding the probability for a system of N independent particles with a total spin S . The total spin S of the system is obviously equal to $s(N_1 - N_2) = sn$, where N_1 and N_2 are the number of particles with spin oriented upward and downward respectively, and $n = N_1 - N_2$. Since $N_1 + N_2 = N$, there correspond to a total spin S , $\frac{1}{2}(N+n)$ particles with a spin directed upward and $\frac{1}{2}(N-n)$ particles with a spin oriented downward. Let us find the number $\Omega(n)$ of independent distributions of $\frac{1}{2}(N+n)$ particles with one spin orientation and $\frac{1}{2}(N-n)$ particles with the other spin orientation for a given total number N of particles.

Each of these distributions leads to one of the equally probable states of the system, so that the probability of the state with spin S which interests us, is according to (16.1), equal to

$$w \sim \Omega(n).$$

N independent particles can be distributed in a given order in $N!$ ways. The mutual permutation of particles with one and the same spin orientation, i.e. $\frac{1}{2}(N+n)$ particles with a spin oriented upward and $\frac{1}{2}(N-n)$ particles with a spin oriented downward, does not change the total spin. The number of such permutations is equal to $[\frac{1}{2}(N+n)]!$ and $[\frac{1}{2}(N-n)]!$. Hence the number of states of the system with a spin S is equal to the number of such independent distributions of N particles between the two states in which $\frac{1}{2}(N+n)$ of the particles are in the first state (with the spin upward) and $\frac{1}{2}(N-n)$ are in the second state (with the spin downward). Correspondingly,

$$w \sim \Omega(n) = \frac{N!}{\left(\frac{N+n}{2}\right)! \left(\frac{N-n}{2}\right)!}.$$

To obtain a more obvious formula use can be made of Stirling's formula (see Appendix IV) and it can be assumed that $n \ll N$. Then, upon taking the logarithm, we have

$$\begin{aligned}\ln w &= \ln N! - \ln \left(\frac{N+n}{2} \right)! - \ln \left(\frac{N-n}{2} \right)! \approx \\ &\approx N \ln \frac{N}{e} - \frac{N+n}{2} \ln \frac{N+n}{2e} - \frac{N-n}{2} \ln \frac{N-n}{2e} \approx N \ln 2 - \frac{n^2}{2N},\end{aligned}$$

or

$$w(n) \sim e^{-n^2/2N}.$$

We have arrived at the probability of a state expressed by a Gaussian distribution. The factor of proportionality can be found from the normalization condition. It is obvious that the most probable state is the state with $n = 0$, i.e. the state in which the number of spins oriented upward is equal to that of spins oriented downward. This state is an analogue of the state of molecular disorder in a gas. The probability distribution has a sharper maximum at the point $n = 0$ the larger the total number of particles in the system.

As we have already stressed, the Gibbs microcanonical distribution, establishing the probability of a given state of a closed system, is of fundamental importance. However, in practice one has much more often to deal not with closed systems, but with subsystems in a reservoir. We shall therefore turn to the consideration of such subsystems.

A subsystem and a reservoir together form a closed system the energy of which (with the above reservation) can be considered to be constant:

$$E = \text{const.}$$

We are interested, however, not in the probability distribution for a complex system but in the probability distribution for the subsystem (for any probability distribution for the reservoir). In order to find this it is necessary to take into account the particular character of the interaction between the subsystem and the reservoir.

As has already been pointed out, this interaction is weak, so that the energy of interaction can be neglected in the total energy balance. Writing the latter in the form

$$E = E_k^{(0)} + \epsilon_i = \text{almost a constant}, \quad (16.2)$$

where $E_k^{(0)}$ is the energy of the reservoir in the k th state, ϵ_i is the energy of

the subsystem in the i th state, and the term "almost a constant" underlines the fact that terms expressing the interaction between the subsystem and the reservoir, as well as between the complex system and the bodies surrounding it, are omitted in the energy conservation law (16.2).

The neglect of the energy of interaction between the system and the reservoir means that we can consider the quasi-closed system and the reservoir to be independent systems for an overwhelmingly large part of the time.

The subsystem can be in any of $\Omega(\epsilon_i)$ states with an energy ϵ_i , while the reservoir can be in any of $\Omega_0(E_k^{(0)})$ states with an energy $E_k^{(0)}$.

If the transitions mentioned do not take the system out of the group of states with an energy ϵ_i and the reservoir correspondingly out of a state with an energy $E_k^{(0)}$, then a change in the state of the subsystem in no way affects the state of the reservoir and, conversely, a change in the state of the reservoir has no effect on the state of the subsystem. On the other hand, by virtue of the energy conservation law (16.2) the energy of the reservoir and that of the subsystem are unambiguously interrelated. If the subsystem has an energy ϵ_i , then the reservoir must have an energy $E_k^{(0)}$.

After these remarks we can now turn to finding the probability that the subsystem be in one of the states with an energy ϵ_i .

From the last remark this probability w_i is equal to the probability that the complex system (subsystem + reservoir) be in such a state that the subsystem has an energy ϵ_i and the reservoir an energy $E_k^{(0)}$. Since w_i is the probability of a given state of a closed system, it is expressed in terms of the number of states according to formula (16.1):

$$w_i \sim \Omega(E) = \Omega(E_k^{(0)} + \epsilon_i). \quad (16.3)$$

On the other hand, the number of states of a closed system consisting of two independent parts is equal to the product of the number of states of the two parts, i.e.

$$\Omega(E_k^{(0)} + \epsilon_i) = \Omega_0(E - \epsilon_i) \Omega(\epsilon_i). \quad (16.4)$$

Here, in the expression for the number of states of the reservoir Ω_0 we have written as the argument the expression $E - \epsilon_i$ on the basis of (16.2).

Substituting the expression (16.4) into (16.3), we find

$$w_i \sim \Omega_0(E - \epsilon_i) \Omega(\epsilon_i). \quad (16.5)$$

The very weak interaction between the system and the surroundings serves

as the cause of transitions of the system from one state into another. Since the dimensions of the reservoir are very large in comparison with those of the system, we can assume that the energy $E_k^{(0)}$ of the reservoir for all values of k is also very large in comparison with the energy of the system. Hence, whatever the changes of the energy of the system may be, the energy of the reservoir can be considered to be almost constant. All different states in which the reservoir is found when the system passes from one energy state into another can be assumed to have one and the same energy.

Because of this we can expand $\Omega_0(E-\epsilon_i)$ in a series in powers of the small quantity ϵ_i and confine ourselves to the first term of the expansion. However, it should be noted that the function $\Omega_0(E-\epsilon_i)$ cannot itself be expanded directly in a power series. Indeed, we know that the number of states is a multiplicative function, while the energy is an additive function. The number of states of a system consisting of independent parts is equal to the product of the number of states of these parts, while the energy is equal to the sum of the corresponding energies. If we expanded $\Omega_0(E-\epsilon_i)$ in a series in powers of the small quantity ϵ_i , then we would obtain the expression

$$\Omega_0(E-\epsilon_i) \approx \Omega_0(E) - \frac{\partial \Omega_0}{\partial E} \epsilon_i, \quad (16.6)$$

which does not possess the required properties. If, for example, we considered two systems with the numbers of states $\Omega_0^{(1)}$ and $\Omega_0^{(2)}$ and energies $[E^{(1)}-\epsilon_i^{(1)}]$ and $[E^{(2)}-\epsilon_i^{(2)}]$ respectively, then the number of states should be equal to $\Omega_0^{(1)}\Omega_0^{(2)}$ and the energy should be equal to $[E^{(1)}-\epsilon_i^{(1)}+E^{(2)}-\epsilon_i^{(2)}]$. However, in multiplying the left-hand sides of the expansion (16.6) the right-hand sides are not multiplied.

Hence before expanding the number of states $\Omega_0(E-\epsilon_i)$ in a series we write it in the form

$$\Omega_0(E-\epsilon_i) = e^{\sigma(E-\epsilon_i)}, \quad (16.7)$$

where $\sigma(E-\epsilon_i)$ is a new function of the argument $(E-\epsilon_i)$. Such a representation is always possible, since the number of states by its very nature is an essentially positive quantity, the values of which are obviously not less than unity.

Writing $\sigma(E-\epsilon_i)$ in the form

$$\sigma(E-\epsilon_i) = \ln \Omega_0(E-\epsilon_i), \quad (16.8)$$

Expanding $\sigma(E - \epsilon_i)$ in a series in powers of the small quantity ϵ_i and confining ourselves to the first term, we have

$$\sigma(E - \epsilon_i) \approx \sigma(E) - \frac{\partial \sigma}{\partial E} \epsilon_i = \sigma - \frac{\epsilon_i}{\theta},$$

where θ stands for the quantity

$$\theta = \left(\frac{\partial E}{\partial \sigma} \right)_{\sigma_i=0}. \quad (16.9)$$

Then for $\Omega_0(E - \epsilon_i)$ we find

$$\Omega_0(E - \epsilon_i) \approx e^{\sigma(E)} e^{-\epsilon_i/\theta}. \quad (16.10)$$

It is easily seen that (16.10) satisfies the requirement of the product of Ω with the addition of the energies of independent systems.

Substituting the expression (16.10) into (16.5), we have

$$w_i = \text{const } e^{-\epsilon_i/\theta} \Omega(\epsilon_i), \quad (16.11)$$

where const stands for the product of the factor of proportionality and the quantity $e^{\sigma(E)}$, which does not depend on the value of ϵ_i or the properties of the subsystem.

Formula (16.11) determines the probability that a certain system, representing a small weakly interacting part of a certain ensemble of arbitrary physical systems, be found in one of $\Omega(\epsilon_i)$ states with an energy lying between ϵ_i and $\epsilon_i + \delta\epsilon_i$, and that the reservoir be found in one of the states with an energy lying between $E - \epsilon_i$ and $E - (\epsilon_i + \delta\epsilon_i)$. Since the state of the reservoir is of no interest, we shall for brevity say that w_i is the probability for the subsystem to be in one of the states with energy ϵ_i .

From the definition of the probability it follows that the following normalization conditions must hold:

$$\sum w_i = 1, \quad (16.12)$$

where the summation is carried out over all possible quantum states of the system.

From the normalization condition and the form of w_i it follows imme-

diately that the coefficient θ , introduced formally, is an essentially positive quantity. Only if this is so does the probability of states of arbitrarily large energies tend to zero, as is to be expected from the very meaning of the notion of physical probability, and as follows formally from the normalization condition. The constant in (16.11) can be found from the normalization condition. Substituting (16.11) into (16.12), we find

$$\text{const} = \frac{1}{\sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i)}.$$

Hence the probability distribution can be put in the final form

$$w_i = \frac{e^{-\epsilon_i/\theta} \Omega(\epsilon_i)}{\sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i)}. \quad (16.13)$$

The distribution (16.13) is the distribution sought, and will serve as the basis of all further discussion. It was first found by Gibbs in 1901 for systems obeying the laws of classical mechanics. This distribution is called the Gibbs distribution or the canonical distribution. The transition from quantum states possessing a discrete set of energy levels to classical systems presents no difficulty and will be made in one of the following paragraphs. The quantity θ figuring in the Gibbs distribution is called the distribution modulus or statistical temperature.

The Gibbs distribution describes the probability distribution of different states of a subsystem representing a small quasi-independent part of an arbitrary system in a state of statistical equilibrium. It should be stressed that, if the system is not in an equilibrium state, then all the foregoing reasoning is invalid. The principle of equal probability of states with a given energy is inapplicable to a non-equilibrium system.

The sum $\sum \exp(-\epsilon_i/\theta) \Omega(\epsilon_i)$ figuring in the denominator of (16.13) will play an important role in what follows. We introduce for it the special notation

$$Z = \sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i), \quad (16.14)$$

and call it the partition function, since all states of the system give a contribution to it. In the literature it is often called the sum over states, or the statistical sum. However, this terminology does not appear to us to be quite apt. By introducing the partition function the Gibbs distribution can be written in the form

$$w_i = Z^{-1} e^{-\epsilon_i/\theta} \Omega(\epsilon_i). \quad (16.15)$$

The Gibbs distribution for any concrete physical system can be considered to be known if the energy levels of the system, i.e. the possible values of the energy ϵ_i , and the degeneracy of the states of the system, i.e. the number of different states $\Omega(\epsilon_i)$ corresponding to a given value of the energy ϵ_i , are known. For a number of systems which will be considered below these physical characteristics can be found.

A remarkable feature of the Gibbs distribution is the fact that the mechanism of interaction of the subsystem with its surrounding in no way figures in it.

By means of the Gibbs distribution one can calculate the mean value of any quantity depending on the state of the system. If $L(\epsilon_i)$ is the value of a certain physical quantity for the states corresponding to an energy ϵ_i , then, according to the general laws for finding a mean value, we can write

$$\begin{aligned} \bar{L} &= \sum L(\epsilon_i) w_i = Z^{-1} \sum L(\epsilon_i) e^{-\epsilon_i/\theta} \Omega(\epsilon_i) = \\ &= \frac{\sum L(\epsilon_i) e^{-\epsilon_i/\theta} \Omega(\epsilon_i)}{\sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i)}. \end{aligned} \quad (16.16)$$

§17. The statistical temperature

Let us first of all consider the properties of the distribution modulus θ which we have already introduced. From its definition it follows that it characterizes the properties of the entire ensemble of systems, i.e. the reservoir, and not the subsystem that we singled out. Indeed, in formula (16.9) there appear only quantities referring to the entire ensemble of subsystems: its energy E and the function σ , whose value $(\partial E/\partial \sigma)$ is taken for $\epsilon_i = 0$, so that $\sigma = \sigma(E)$. Hence the modulus θ always refers to a macroscopic system and is a function of the state of this system. When the state, and in particular the energy of the entire system changes, the distribution modulus θ changes. Since the function σ , determined from formula (16.8) and representing the logarithm of the number of states with a given energy, is a

single-valued function of the state (energy) of the system, θ is also a single-valued function of the energy of state of the system.

Further, the distribution modulus θ is an essentially positive quantity. Indeed, if the energy ϵ_i can assume any arbitrarily large value, then the probability of a state with given energy ϵ_i must decrease with increasing energy. If that were not so, the normalization condition (16.12) could not be fulfilled.

Thus, θ can refer only to a macroscopic system and is essentially a positive single-valued function of its state. We shall show that the distribution modulus θ is a characteristic of the equilibrium state in the system. For this we shall consider two subsystems belonging to different systems, having distribution moduli θ_1 and θ_2 . Each of the subsystems will be assumed to be in a state of statistical equilibrium, so that the probabilities of their states are determined by formula (16.11):

$$w_1 = A_1 e^{-\epsilon_1/\theta_1} \Omega_1, \quad w_2 = A_2 e^{-\epsilon_2/\theta_2} \Omega_2.$$

We assume that both systems come into a weak interaction, so that an energy exchange can take place between them. The two interacting subsystems can be considered as one unified subsystem. If the latter turns out to be in a state of statistical equilibrium, then the probability distribution of its states must also be described by a law of the form

$$w = A e^{-\epsilon/\theta} \Omega. \quad (17.1)$$

On the other hand, since the interaction is weak, the interaction energy can be disregarded and each of the subsystems can be assumed to be quasi-independent. Then to find the probability distribution of the complex system use can be made of the theorem of multiplication, and one can write

$$w = w_1 w_2 = A_1 A_2 e^{-\epsilon_1/\theta_1} e^{-\epsilon_2/\theta_2} \Omega_1 \Omega_2. \quad (17.2)$$

In order that the distribution (17.2) may be identical with (17.1), it is necessary that

$$\theta_1 = \theta_2 = \theta.$$

Thus, if two equilibrium subsystems with equal moduli $\theta_1 = \theta_2$ are brought into interaction, then a unified equilibrium system with the same modulus $\theta = \theta_1 = \theta_2$ will be obtained. If θ_1 were different from θ_2 , then in establish-

ing the interaction a system would arise with a probability distribution expressed by formula (17.2). This distribution is not the Gibbs distribution for a system with energy $\epsilon = \epsilon_1 + \epsilon_2$. Hence the system formed when $\theta_1 \neq \theta_2$ will not be in an equilibrium state. The equilibrium state is not violated when an interaction is established between subsystems, if their moduli θ_1 and θ_2 are equal to each other, and is violated if $\theta_1 \neq \theta_2$.

That is why the quantity θ is called the statistical temperature. In the case where the subsystem contains such a large number of particles that it can be considered as macroscopic, one can also speak of its proper statistical temperature. Its temperature is determined from the condition of equilibrium of the subsystem and the reservoir and, consequently, is equal to the temperature of the latter. Therefore, θ can, for brevity, be called the temperature of the system.

It goes without saying that, if a quasi-closed subsystem contains an insufficiently large number of particles, then the notion of its temperature becomes approximate and, in the case of a subsystem represented by an individual molecule of an ideal gas, it completely loses any meaning.

The value of the statistical temperature is determined by formula (16.9) and depends on the energy of the system. It is, in the general case, impossible to find the form of this dependence, since it is determined by the particular properties of the system. However, in practice one is interested not in the dependence of θ on E but in the reverse, the dependence of the energy on the temperature $E = E(\theta)$. In what follows we shall see that the energy is a monotonic function of temperature. We shall find the concrete form of the dependence of the energy on the temperature θ for certain simple systems (gas, ideal crystal and so on).

§18. The properties of the Gibbs distribution and statistical equilibrium

The Gibbs distribution characterizes the probability distribution of different states of a quasi-closed system. The conditions for the Gibbs distribution to be applicable are:

- (1) the presence of a certain macroscopic system constituting the surroundings (reservoir) of the system considered.
- (2) the presence of a weak interaction between the system and the reservoir.

Other properties of the system are completely arbitrary.

The Gibbs distribution, as well as the Maxwell energy distribution, has a maximum at a certain energy value. At first sight the existence of this maxi-

num is not obvious: the exponentially decreasing factor $\exp(-\epsilon_i/\theta)$ figures in the Gibbs distribution.

However, it should be recalled that the number of states with given energy $\Omega(\epsilon_i)$ increases rapidly with the energy of the system. The more particles the system contains, the more states $\Omega(\epsilon_i)$ correspond to a given energy interval $\epsilon_i, \epsilon_i + \delta\epsilon_i$. Hence the increase of $\Omega(\epsilon_i)$ with energy proceeds the more rapidly the more particles there are in the system. As will, for example, be shown in § 20, if the subsystem is a gas consisting of N independent monatomic molecules, confined in a container with a constant temperature (reservoir), then $\Omega(\epsilon) \sim e^{-\frac{3}{2}N}$.

The product of two functions — the one decreasing rapidly and the other increasing with energy — leads to the appearance of a sharp maximum in the Gibbs distribution. This maximum is sharper, the more rapidly $\Omega(\epsilon_i)$ increases, i.e. the more particles there are in the system. We shall see in the same example that if the system is macroscopic, so that it contains a very large number of particles, then the width of the maximum is negligible. It is so sharp that it is impossible to represent the Gibbs distribution graphically without distorting the scale. This means that the probability for the system to be in states with an energy differing appreciably from the energy $\epsilon_{m,p} = \epsilon_{\max}$, corresponding to the maximum of the Gibbs distribution, is negligibly small (fig. III.11). The system spends an overwhelmingly large part of the time of observation in states with an energy very close to the latter. The state corresponding to the maximum of the Gibbs distribution is the most probable. The most probable state will give the basic contribution to the mean value of the quantities characterizing the system (for example, the energy). This

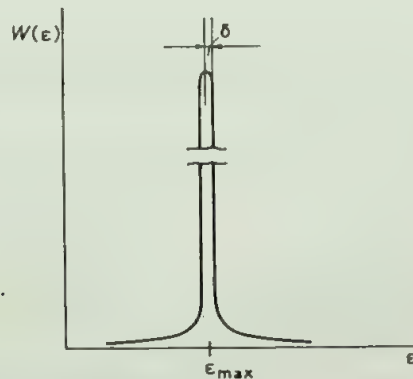


Fig. III.11

follows from the very definition of the idea of mean value: each state gives to the mean value a contribution proportional to its probability.

Hence in the case of a macroscopic system the partition function Z can be written in the following form:

$$Z = \sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i) \approx e^{-\epsilon_{m.p.}/\theta} \Omega(\epsilon_{m.p.}) \approx e^{-\bar{\epsilon}/\theta} \Omega(\bar{\epsilon}), \quad (18.1)$$

where only one term, the largest, referring to the most probable energy, is retained in the sum over states. Here we have made use of the approximate equality of the most probable energy and the mean energy in a macroscopic system.

Analogously, for the mean value of any quantity L we can write

$$\bar{L} = \sum L(\epsilon_i) w(\epsilon_i) \approx L(\epsilon_{m.p.}) \approx L(\bar{\epsilon}), \quad (18.2)$$

i.e. the state with $\epsilon = \epsilon_{m.p.}$ is realized with a probability $w(\epsilon_{m.p.}) \approx 1$, while probability of other states $\epsilon \neq \epsilon_{m.p.}$ is close to zero. The mean value of all quantities will be close to their most probable value. This refers, in particular, also to the energy $\bar{\epsilon} \approx \epsilon_{m.p.}$ of the system. This result is in a complete agreement with the general conclusions arrived at in §5 on the properties of systems containing a large number of particles. The true values of all quantities are close to their mean values, and the latter are close to the most probable values.

The presence of a sharp maximum in the Gibbs distribution is a concrete manifestation of the general properties of systems with a large number of particles considered in §5.

Systems which are in such a state that the true values of the quantities characterizing them are close to mean values are called systems in a state of statistical equilibrium.

Thus we see that every macroscopic quasi-closed system described by the Gibbs distribution is in a state of statistical equilibrium during most of the time of observation.

§19. Transition to classical statistics

In most cases we shall have to deal with systems the energy levels of which are so close to each other that they can be considered to be distributed con-

tinuously. Then the set of discrete values of the energy levels $\epsilon_1, \epsilon_2, \dots, \epsilon_i, \dots$ can be replaced by a continuous function ϵ . In other words, we shall pass over from the quantal description of the system to the quasi-classical one in the sense explained in § 1.

From the Gibbs distribution it follows that, in order to replace a discontinuous function $\exp(-\epsilon_i/\theta)$ by a smooth function $\exp(-\epsilon/\theta)$, it is necessary that the magnitude of the steps, i.e. of the spacing between the levels $\Delta\epsilon_i = \epsilon_{i+1} - \epsilon_i$, should be small in comparison with the value of θ . Thus, other things being equal, the transition to quasi-classical statistics should take place in the range of high temperature. We shall frequently return to this statement in what follows. The application of the classical approach is of great importance and is often encountered in practice, since, in ordinary conditions, the energy spectrum of every macroscopic system is almost continuous.

The statistical physics of systems obeying the laws of classical mechanics is called classical statistics.

In the classical approximation we have to replace a discrete set of probabilities for different states by a continuous distribution. In the quasi-classical approximation the state of a system of N particles having $3N$ degrees of freedom is determined by the value of the coordinates q_1, q_2, \dots, q_{3N} and p_1, p_2, \dots, p_{3N} . The energy of the system $\epsilon(p, q)$ is expressed as a continuous function of all the coordinates and momenta. Since energy in the classical approximation can be considered to be a continuous function, the probability distribution of different states of the system is also expressed by a continuous function. Namely, in the classical approximation one can give the probability $d\omega$ for a system to have an energy lying between $\epsilon(p, q)$ and $\epsilon(p, q) + d\epsilon(p, q)$. According to (1.26), there corresponds to an energy lying in this interval a number of states $d\Omega = h^{-3N}(\partial\Gamma/\partial\epsilon)d\epsilon$, where N is the number of particles in the subsystem.

In the classical approximation the Gibbs distribution can be written in the form

$$d\omega = \frac{e^{-\epsilon(p, q)/\theta}}{Z} \frac{\partial\Gamma}{\partial\epsilon} \frac{d\epsilon}{h^{3N}}. \quad (19.1)$$

Here according to (16.14), the partition function of the system can be written in the form

$$Z = \frac{1}{h^{3N}} \int e^{-\epsilon(p,q)/\theta} \frac{\partial \Gamma}{\partial \epsilon} d\epsilon. \quad (19.2)$$

The difference between (19.2) and (16.14) lies in the fact that in the classical formula for Z an integration is carried out instead of a summation. Z is often called the statistical integral. The integration in (19.2) is carried out over the entire phase space available for the system, i.e. over all allowed values of the coordinates and momenta. Which values of the coordinates and momenta are allowed depends on the particular properties of the system and on the conditions in which it is found.

Substituting (19.2) into (19.1), we have

$$dw = \frac{e^{-\epsilon(p,q)/\theta} (\partial \Gamma / \partial \epsilon) d\epsilon}{\int e^{-\epsilon(p,q)/\theta} (\partial \Gamma / \partial \epsilon) d\epsilon} = \frac{e^{-\epsilon(p,q)/\theta} d\Gamma}{\int e^{-\epsilon(p,q)/\theta} d\Gamma}. \quad (19.3)$$

From this it is clear that Planck's constant drops out of the Gibbs classical distribution dw , as was to be expected. The Gibbs classical distribution is often written in the form

$$dw = \rho(p, q) d\Gamma, \quad (19.4)$$

where $\rho(p, q)$ is the normalized probability density:

$$\rho(p, q) d\Gamma = \frac{e^{-\epsilon(p,q)/\theta} d\Gamma}{\int e^{-\epsilon(p,q)/\theta} d\Gamma}; \quad \int \rho(p, q) d\Gamma = 1. \quad (19.5)$$

Here dw represents the probability of a given state of the system, i.e. the probability for the representative point of the system to be in a given element of phase space.

In other words, dw represents the probability for the system to be in a state in which its momenta and coordinates lie in the intervals $p_1, p_1 + dp_1; \dots, p_{3N}, p_{3N} + dp_{3N}$ and $q_1, q_1 + dq_1; \dots, q_{3N}, q_{3N} + dq_{3N}$.

Consider, in particular, the case when the quasi-closed subsystem is an individual molecule in an ideal gas. Then the energy of the subsystem $\epsilon(p, q)$ represents the energy of this molecule. For $N = 1$ our subsystem will have three degrees of freedom. Correspondingly, its phase space will have six dimensions. An element $d\Gamma$ of the phase volume will have the form

$$d\Gamma = dp_x dp_y dp_z dx dy dz$$

or, in spherical coordinates for the momenta, replacing for conciseness $dx dy dz$ by a volume element dV ,

$$d\Gamma = p^2 dp \sin \theta d\theta d\varphi dV.$$

If there are no external fields of force the energy of the molecule reduces to its kinetic energy:

$$\epsilon(p, q) = p^2/2m,$$

and depends neither on the direction of its motion (angles θ and φ) nor on its position in the container. Hence there corresponds to an energy lying in the interval between ϵ and $\epsilon + d\epsilon$ a number of states equal to

$$d\Omega = \frac{p^2 dp}{d\epsilon h^3} \int dV \int \sin \theta d\theta d\varphi d\epsilon = \frac{4\pi V}{h^3} p^2 \frac{dp}{d\epsilon} d\epsilon.$$

The calculation of $dp/d\epsilon$ gives.

$$d\Omega = \frac{4\pi V m^{\frac{3}{2}} (2\epsilon)^{\frac{1}{2}}}{h^3} d\epsilon.$$

Thus, the Gibbs distribution for one molecule has the form

$$dw = \frac{4\pi V m^{\frac{3}{2}}}{zh^3} e^{-\epsilon/\theta} (2\epsilon)^{\frac{1}{2}} d\epsilon, \quad (19.6)$$

where z is the partition function of an individual molecule. Comparison of (19.6) with the Maxwell energy distribution (9.5) convinces us of their identity, provided the statistical temperature θ is identified with the quantity kT . It should be stressed that the absolute temperature, appearing in the Maxwell distribution, refers not to an individual molecule (the subsystem) but to the entire gas (the reservoir). In § 26 we shall show that this relationship between θ and T is of a general character. At first sight it may seem that the normalization constant in the Maxwell distribution differs from that in (19.3). In particular, it does not contain the Planck constant h . In reality, however, this is not so. In order to convince ourselves of this, we write the explicit expression for z :

$$\begin{aligned}
 z &= \int e^{-\epsilon/\theta} \frac{d\Gamma}{h^3} = \frac{4\pi V m^{\frac{3}{2}}}{h^3} \int_0^\infty e^{-\epsilon/\theta} (2\epsilon)^{\frac{1}{2}} d\epsilon = \\
 &= \frac{4\pi V m^{\frac{3}{2}}}{h^3} \left(\frac{\pi \theta^3}{2} \right)^{\frac{1}{2}} = \left(\frac{2\pi m \theta}{h^2} \right)^{\frac{3}{2}} V.
 \end{aligned} \tag{19.7}$$

Hence

$$dw = \frac{2}{(\pi \theta^3)^{\frac{1}{2}}} e^{-\epsilon/\theta} \epsilon^{\frac{1}{2}} d\epsilon. \tag{19.8}$$

Thus, the Planck constant vanishes from the distribution, and the constants in (19.8) and (9.5) are the same.

As we have stressed in the preceding paragraph, the Gibbs distribution has a very sharp maximum for a certain value of energy. At first sight this statement contradicts the broad maximum in the Maxwell distribution. However, it should be borne in mind that the sharp maximum in the Gibbs distribution arises as a result of the competition of the exponentially decreasing factor $\exp(-\epsilon/\theta)$ and the increasing factor $\Omega(\epsilon)$. The latter increases as $\epsilon^{\frac{3}{2}N}$, or as $\epsilon^{\frac{3}{2}}$ in the case when $N = 1$. Hence for $N \gg 1$ the quantity $dw/d\epsilon$ varies rapidly and a sharp maximum arises, while for $N = 1$ it increases relatively slowly and the maximum in the distribution turns out to be broad.

If the quasi-classical subsystem contains a very large number of particles, then the integral over states, figuring in formula (19.2), has a very sharp maximum for an energy value $\epsilon_{\max} \approx \epsilon$, i.e. in the region of states corresponding to statistical equilibrium of the system.

In this case, analogously to (18.1), we can write

$$Z \approx e^{-\epsilon_{m.p}/\theta} \frac{\Delta\Gamma}{h^{3N}} = e^{-\bar{\epsilon}/\theta} \frac{\Delta\Gamma}{h^{3N}}, \tag{19.9}$$

where $\Delta\Gamma$ is the volume of that region of the phase space which corresponds to a state of statistical equilibrium, i.e. to $\epsilon \approx \epsilon_{m.p.}$. It is obvious that the number of states corresponding to statistical equilibrium of the system is equal to

$$\Omega(\epsilon_{m.p.}) = \frac{\Delta\Gamma}{h^{3N}}. \tag{19.10}$$

§ 20. The monatomic gas as a whole

The properties of the Gibbs distribution, described in § 18 and 19, can be shown most clearly by a concrete example.

Let us consider as a unique quasi-closed system, a gas as a whole, confined in a container of volume V . If the walls of the container are impenetrable to but can exchange energy with the gas molecules, then the walls of the container and the bodies surrounding the container form a reservoir. The entire container with the gas can be characterized by a definite temperature θ , equal to the temperature of the surrounding bodies. It can be assumed that the size of the latter and their energy are very large in comparison with the energy of the gas.

We see that all the conditions for the applicability of the Gibbs distribution to a gas as a whole are present, and that this distribution can be written for the gas as a whole. We assume that the gas is monatomic, and that there are no external fields of force. Then the energy of the gas is equal to the sum of the kinetic energies of all the particles constituting it. This is given by the classical expression and is a continuous variable. Let the gas contain N molecules of mass m . The state of the system is completely characterized by giving the coordinates and momenta of all molecules $q_1, q_2, \dots, q_{3N}, p_1, p_2, \dots, p_{3N}$. The phase space of the system has $6N$ dimensions. An element $d\Gamma$ of the phase space is equal to the product of the differentials of all the momenta and coordinates:

$$d\Gamma = dp_1 \dots dp_{3N} dq_1 \dots dq_{3N} . \quad (20.1)$$

The energy of the system depends only on the momenta of the molecules and can be written in the form

$$\epsilon(p, q) = \frac{1}{2m} (p_1^2 + p_2^2 + \dots + p_{3N}^2) . \quad (20.2)$$

In order to write down the Gibbs distribution it is necessary to find an expression for the number of states corresponding to the energy of the system lying between ϵ and $\epsilon + \delta\epsilon$. According to the general formula (1.26') we have

$$d\Omega = \frac{1}{h^{3N}} \frac{\partial \Gamma}{\partial \epsilon} d\epsilon . \quad (20.3)$$

The Gibbs distribution for the gas as a whole has the form

$$dw = \frac{1}{h^{3N}Z} \exp \left[-\frac{p_1^2 + p_2^2 + \dots + p_{3N}^2}{2m\theta} \right] \frac{\partial \Gamma}{\partial \epsilon} d\epsilon. \quad (20.4)$$

Let us find the value of $\partial \Gamma / \partial \epsilon$. The volume of the part of phase space in which the energy of the gas does not exceed ϵ is, by definition, equal to

$$\Gamma = \int dp_1 \dots dp_{3N} dq_1 \dots dq_{3N}. \quad (20.5)$$

In formula (20.5) the integration range is determined in such a way that the condition

$$\frac{p_1^2 + p_2^2 + \dots + p_{3N}^2}{2m} \leq \epsilon(p, q) \quad (20.6)$$

is fulfilled. This condition does not involve the coordinates of the molecules, with respect to which one can integrate directly. This gives

$$\Gamma = V^N \int dp_1 \dots dp_{3N}, \quad (20.7)$$

where $V = \int dq_1 dq_2 dq_3$ is the entire volume of the gas.

Formula (20.6) determines, from the geometrical point of view, in the space of $3N$ dimensions a sphere whose radius is equal to $R = (2m\epsilon)^{1/2}$. Then the integral in (20.7) represents the volume of this sphere. The dependence of the volume of the sphere of $3N$ dimensions on its radius can be found from the consideration of dimensionality. Namely, it must be proportional to the radius to a power equal to the number of dimensions. In the three-dimensional space it is proportional to R^3 , and in the $3N$ -dimensional space it is proportional to R^{3N} . Hence (20.7) can be written in the form

$$\Gamma = \text{const } V^N R^{3N} = \text{const } V^N \epsilon^{\frac{3}{2}N}. \quad (20.8)$$

Differentiating (20.8) we have

$$\frac{\partial \Gamma}{\partial \epsilon} = \text{const } V^N \epsilon^{\frac{3}{2}N-1}. \quad (20.9)$$

The value of the constant in (20.9) is of no particular interest, since it will be cancelled with the same constant arising in calculating Z . Hence, from (20.4) and (20.9) we finally have,

$$dw = \frac{\text{const}}{h^{3N}Z} e^{-\epsilon/\theta} \epsilon^{\frac{3}{2}N-1} V^N d\epsilon. \quad (20.10)$$

The Gibbs distribution function for a system with a large number of particles N has a very sharp maximum, since the factor $\epsilon^{\frac{3}{2}N-1} \approx \epsilon^{\frac{3}{2}N}$ increases very rapidly with increasing ϵ , whereas the factor $\exp(-\epsilon/\theta)$ decreases sharply. We shall find the position, width and height of this maximum.

The maximum of the expression (20.10) is at a point determined by the condition

$$\frac{d}{d\epsilon} e^{-\epsilon/\theta} \epsilon^{\frac{3}{2}N-1} = -\frac{e^{-\epsilon/\theta} \epsilon^{\frac{3}{2}N-1}}{\theta} + \left(\frac{3}{2}N-1\right) e^{-\epsilon/\theta} \epsilon^{\frac{3}{2}N-2} = 0. \quad (20.11)$$

Hence we find that the condition for the maximum reads

$$-\frac{\epsilon_{\text{max}}}{\theta} + \left(\frac{3}{2}N-1\right) = 0$$

or

$$\epsilon_{\text{max}} = \epsilon_{\text{m.p}} = \left(\frac{3}{2}N-1\right) \theta,$$

where ϵ_{max} is the energy at the maximum. Since the number of particles N is very large, unity can be neglected in comparison with $3N$, so that

$$\epsilon_{\text{max}} = \epsilon_{\text{m.p}} \approx \frac{3}{2}N\theta. \quad (20.12)$$

It is easy to show that the quantity $\frac{3}{2}N\theta$ represents the mean energy of the entire gas. By definition,

$$\begin{aligned} \bar{\epsilon} &= \int \epsilon dw = \frac{\text{const}}{h^{3N}Z} V^N \int e^{-\epsilon/\theta} \epsilon^{\frac{3}{2}N-1} \epsilon d\epsilon \approx \\ &= \frac{\text{const } V^N}{h^{3N}} \frac{\int_0^\infty e^{-\epsilon/\theta} \epsilon^{\frac{3}{2}N} d\epsilon}{\frac{\text{const } V^N}{h^{3N}} \int_0^\infty e^{-\epsilon/\theta} \epsilon^{\frac{3}{2}N-1} d\epsilon} = \frac{\partial}{\partial(-\theta^{-1})} \ln \int_0^\infty e^{-\epsilon/\theta} \epsilon^{\frac{3}{2}N-1} d\epsilon. \end{aligned} \quad (20.13)$$

The integral in (20.13) is calculated in Appendix IV. This calculation leads to the relation

$$\bar{\epsilon} = \frac{3}{2}N\theta. \quad (20.14)$$

Comparing the expressions (20.12) and (20.14), we see that the most probable energy lies very close to the mean energy. If N is sufficiently large, then these energies can be identified with each other to a high degree of accuracy. Thus, the subsystem (ideal gas) during an overwhelmingly large part of the time is in the state in which its energy is equal to the mean energy $\bar{\epsilon}$. This property is not possessed by a subsystem containing few particles. For example, for one molecule the difference between the mean and the most probable energy is relatively large.

In order to picture how sharp the maximum in the Gibbs distribution is, i.e. how often the subsystem can get into a state with an energy different from the most probable one ϵ_{\max} , we shall find the form of the distribution function near the maximum. In the vicinity of the maximum, when the difference $\epsilon - \epsilon_{\max}$ is small, the distribution function can be expanded in a series in powers of $\epsilon - \epsilon_{\max}$ and we can restrict ourselves to the first terms of the expansion. If the distribution function (ignoring the immaterial constant) is denoted by f , then

$$f = e^{-\epsilon/\theta} \epsilon^{\frac{3}{2}N-1} = e^{-\epsilon/\theta + (\frac{3}{2}N-1) \ln \epsilon} = e^{\varphi(\epsilon)},$$

where

$$\varphi(\epsilon) = -\frac{\epsilon}{\theta} + \left(\frac{3}{2}N-1\right) \ln \epsilon.$$

Since at the point $\epsilon = \epsilon_{\max}$ the distribution function f and, therefore, also the function φ has a maximum, for $\varphi(\epsilon)$ near this maximum one can write the expansion

$$\begin{aligned} \varphi(\epsilon) &\approx \varphi(\epsilon_{\max}) + \left(\frac{d\varphi}{d\epsilon}\right)_{\epsilon=\epsilon_{\max}} (\epsilon - \epsilon_{\max}) + \\ &+ \frac{1}{2} \left(\frac{d^2\varphi}{d\epsilon^2}\right)_{\epsilon=\epsilon_{\max}} (\epsilon - \epsilon_{\max})^2 + \dots \approx \\ &\approx \varphi(\epsilon_{\max}) + \frac{1}{2} \left(\frac{d^2\varphi}{d\epsilon^2}\right)_{\epsilon=\epsilon_{\max}} (\epsilon - \epsilon_{\max})^2. \end{aligned}$$

It is easy to see that

$$\left(\frac{d^2\varphi}{d\epsilon^2}\right)_{\epsilon=\epsilon_{\max}} = -\left(\frac{3}{2}N-1\right) \frac{1}{\epsilon_{\max}^2}$$

and, consequently,

$$\begin{aligned} f &\approx \exp[\varphi(\epsilon_{\max})] \exp\left[-\left(\frac{3}{2}N-1\right) \frac{(\epsilon-\epsilon_{\max})^2}{2\epsilon_{\max}^2}\right] = \\ &= \exp[-\epsilon_{\max}/\theta] \epsilon_{\max}^{\frac{3}{2}N-1} \times \exp\left[-\left(\frac{3}{2}N-1\right) \frac{(\epsilon-\epsilon_{\max})^2}{2\epsilon_{\max}^2}\right]. \end{aligned}$$

Thus, the probability distribution near the maximum has the form

$$\begin{aligned} dw &= \text{const} \frac{1}{h^{3N}Z} \exp[-\epsilon_{\max}/\theta] \epsilon_{\max}^{\frac{3}{2}N-1} \times \\ &\times \exp\left[-\frac{(\frac{3}{2}N-1)(\epsilon-\epsilon_{\max})^2}{2\epsilon_{\max}^2}\right] d\epsilon. \end{aligned} \quad (20.15)$$

The dependence of the probability distribution on the distance from the maximum $(\epsilon-\epsilon_{\max})$ is characterized by the second exponential factor in (20.15). It represents a symmetric function of the type

$$\exp\left[-\frac{(\epsilon-\epsilon_{\max})^2}{2\delta^2}\right], \quad \text{where} \quad \delta = \frac{\epsilon_{\max}}{(\frac{3}{2}N-1)^{\frac{1}{2}}}.$$

The quantity δ represents the width of the maximum. At the value $(\epsilon-\epsilon_{\max}) = \delta$ the distribution function is e times smaller than at the maximum. The relative width of the maximum is equal to

$$\frac{\delta}{\epsilon_{\max}} \approx \frac{\delta}{\bar{\epsilon}} \approx \frac{1}{(\frac{3}{2}N-1)^{\frac{1}{2}}} \approx \left(\frac{2}{3N}\right)^{\frac{1}{2}} \quad (20.16)$$

For values of N corresponding to the number of molecules in a macroscopic volume ($N \approx 10^{19}$), the width of the maximum of the Gibbs distribution turns out to be very small. This means that the Gibbs distribution has a very sharp maximum at ϵ_{\max} . The gas is, with a high probability, in the state in which its energy is equal to the mean energy $\bar{\epsilon}$. The probability that 1 cm³ of

the gas will be found in a state with an energy differing from $\bar{\epsilon}$, for example an ϵ equal to 99% of $\bar{\epsilon}$, can easily be determined from formula (20.15) or (20.16). Its ratio to the probability of the state $\epsilon = \bar{\epsilon}$ is as

$$1 : \exp \left[\frac{1}{2} \left(\frac{3}{2} N - 1 \right) (0.99 - 1)^2 \right] = 1 : e^{10^{15}}.$$

Thus, an appreciable deviation of the energy of the gas from the mean value does not in practice occur in a gas containing a large number of particles. This conclusion is in complete agreement with what was said in the previous paragraph, as well as with the general theorem of §5. Comparing the relative width of the maximum of (20.16) with the determination of the relative energy fluctuation of §5, we convince ourselves of their complete equivalence.

Statistical and Phenomenological Thermodynamics

§21. The internal energy of a macroscopic system. The first and second laws of thermodynamics

Turning from the apparatus of Gibbs statistics, we now consider the formulation of a theory of the thermal properties of matter. Before proceeding to carry out this programme, it is necessary to dwell briefly on the history of the development of the theory of heat.

The development of technology and the extensive use of heat engines in the first half of the 19th century demanded the development of a theory of thermal processes. However, the ideas of the nature of heat were still very vague. The physics of the first half of the 19th century was still very far from the construction of a theory of thermal processes on the basis of molecular concepts. Hence the development of the theory proceeded in a very unusual way.

Joule's experimental establishment of the mechanical equivalent of heat and the failure of all attempts to create a perpetual motion machine (perpetuum mobile), by means of which it would be possible to obtain useful work without any changes in surrounding bodies, allowed one to postulate a general principle called the first law of thermodynamics. The first law of thermodynamics represents a particular case of the energy conservation law as applied to thermal processes.

If the amount of heat absorbed by a system and expressed in mechanical units is equal to δQ , then the first law of thermodynamics reads:

$$\delta Q = -\delta W + \delta E, \quad (21.1)$$

where $\{-\delta W\}$ is the mechanical work done by external forces on the system absorbing heat. The difference $\delta Q - (-\delta W)$ between the heat absorbed and the work done represents the part of the heat used in changing the internal state of the system.

The quantity E , representing a function of the internal state of the system, is called the internal energy. For a so-called cyclic process, in which the system after all changes comes back into its initial state, the algebraic sum of all heat absorbed and work done is equal to zero. This means that in a cyclic process the system obtains the same amount of heat from without as the mechanical work it gives. Whence it follows that the change in the internal energy of a system in a cyclic process is equal to zero, $\oint dE = 0$. This equality means that the internal energy is a single-valued function of the state of the system. Thus, the relation (21.1) expresses the energy conservation law. It is usually written in the form

$$dE = \delta Q + \delta W.$$

The second basic proposition of phenomenological thermodynamics, called the second law of thermodynamics, also represents a generalization of the results of numerous experimental data. The second law of thermodynamics says that it is impossible to draw heat systematically from a system and to transform it into work without some other changes occurring simultaneously in the system or in the bodies surrounding it.

A machine which, drawing heat from a body, could transform it systematically into work was called a perpetual motion machine of the second kind. It is clear that, if such a machine could be constructed, the bodies surrounding us, for example oceans, could serve as a practically inexhaustible reservoir of work. However, all attempts to construct such a machine failed.

Thus, the second law of thermodynamics, like the first, rested upon numerous reliable experimental data. In what follows, it will be shown how one can pass over from the above qualitative formulation of the second law to its quantitative formulation. It turned out that, based upon the mathematical formulation of the first and second laws of thermodynamics, it was possible to construct a phenomenological theory of thermal processes, called thermo-

dynamics. All deductions of thermodynamics had the same degree of reliability as the first and second laws on which they were based, which made them indisputable.

The reliability and general character of the deductions represent the most important merit of thermodynamical methods of investigation. Their shortcomings lie in the fact that they do not disclose the physical, molecular basis of thermal processes. Hence the construction of a molecular theory of heat and the elucidation of the molecular basis of thermodynamical ideas appeared to be the most important stage in the development of the theory of heat and physics as a whole.

At present thermodynamics and the molecular theory of thermal processes (statistical thermodynamics) make a unified whole.

In what follows we shall be able to convince ourselves by concrete examples that phenomenological and statistical thermodynamics do not contradict, but supplement each other.

We shall base the molecular theory of the thermal properties of matter on the following very natural supposition:

“The internal energy of a macroscopic body is identical with the mean energy \bar{E} calculated according to the laws of statistical physics”.

In the following we shall consider the thermal properties of macroscopic systems which contain a very large number of particles and which are in a state of statistical equilibrium. Since in a system containing a large number of particles and in a state of statistical equilibrium the mean energy \bar{E} is practically the same as its actual energy, this assumption can be formulated differently:

“The internal energy of any macroscopic body represents the energy of thermal motion of the molecules constituting the body”.

It should be noted that at present this assumption is so well substantiated experimentally and theoretically that the term “assumption” appears to be inadequate.

We do not, however, consider it superfluous to stress that the identification of the mean energy \bar{E} of motion of the molecules with the thermodynamic energy E is the basis of our further exposition. All other statements, having a less obvious character, for example the statistical treatment of the second law of thermodynamics, which we shall analyse in later paragraphs, do not need any new assumptions or references to experiment for their substantiation, but appear to be a direct consequence of this unique assumption.

In order actually to calculate the mean energy of the system we have to make use of the general rule of § 16 (eq. (16.16)) which, as applied to the energy, reads

$$\bar{\epsilon} = \frac{\sum \epsilon_i e^{-\epsilon_i/\theta} \Omega(\epsilon_i)}{\sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i)},$$

where the summation is carried out over all energy levels of the system.

The expression for the mean energy $\bar{\epsilon}$ can be rewritten in a more compact form. That is, from the obvious identity

$$\frac{\partial}{\partial(-\theta^{-1})} \sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i) = \frac{\partial Z}{\partial(-\theta^{-1})} = \sum e^{-\epsilon_i/\theta} \epsilon_i \Omega(\epsilon_i)$$

it follows that $\bar{\epsilon}$ can be written in the form

$$\bar{\epsilon} = \frac{\partial}{\partial(-\theta^{-1})} \ln Z = \theta^2 \frac{\partial \ln Z}{\partial \theta}. \quad (21.2)$$

From formula (21.2) it follows that to find the mean energy of the system it is sufficient to know its partition function Z . Because of the assumption of the identity of the mean and thermodynamic energies of a system we shall always write

$$E = \frac{\partial}{\partial(-\theta^{-1})} \ln Z = \theta^2 \frac{\partial \ln Z}{\partial \theta}. \quad (21.3)$$

It follows from the formulae given that the state of a macroscopic system, in particular its internal energy, depends on the temperature θ of the reservoir. In a state of statistical equilibrium the temperature of the system is equal to the temperature of its surroundings (reservoir), so that one can speak about the dependence of the energy of the body on its proper temperature.

The internal energy of a macroscopic system possesses the important property of additivity: the energy of a complex system is equal to the sum of the energies of its macroscopic parts. This statement is, of course, of approximate character. It assumes that the energy of interaction between the parts can be neglected. In the case of macroscopic parts it usually can be neglected, since it has the character of a surface energy (see, for instance, §65).

§ 22. Work and pressure

The state of a body in statistical equilibrium depends on the external conditions as well as the temperature, which are determined by the magnitude of external fields acting on the body.

According to what was said at the beginning of § 8, the volume of a body is also determined by fields of force acting on the surface of the body; the walls of the container represent such a field of force as shown in fig. III.5.

External conditions can be characterized by certain quantities called external parameters. These external parameters of a system are determined by the fields acting on the body or by the position of bodies surrounding it.

Imagine, for example, that our system is a gas confined in a container with a movable cover (a piston). Then the state of the system depends on the position of the piston. This position is an external parameter, since the value of the coordinate of the piston does not depend on the nature and properties of the system in the container. As a second example one can cite a system in an external field of force. If an arbitrary system is in such an external field of force, then its particles possess a certain potential energy. Hence the energy levels will depend on the properties of the field. In a uniform field this dependence is determined only by the position of the system in the field. In this case the position of the system serves as an external parameter.

Thus, the energy levels of a system, in general, depend not only on the properties of the system itself but also on the values of external parameters, the set of which we shall denote by λ . To stress this, we shall sometimes write $\epsilon_i(\lambda)$. However, it should not be forgotten that the values ϵ_i depend not only on λ but also on the properties of the system itself.

Consider the change $\delta\epsilon_i$ in the energy of the system for an infinitely small change $\delta\lambda$ of its external parameters. We confine ourselves from the beginning to such a change of external parameters for which the probability distribution of different states remains unchanged. This means that, when the external parameters are changed, a transition of the system from one state into another does not occur (see below).

We then have

$$\delta\epsilon_i = \frac{\partial\epsilon_i}{\partial\lambda} \delta\lambda. \quad (22.1)$$

The quantity $\partial\epsilon_i/\partial\lambda$ (taken with the opposite sign) can be considered as a generalized force acting on the system. We denote it by $(-f_i)$. Then (22.1) can be written in the form

$$\delta \epsilon_i = -f_i \delta \lambda. \quad (22.2)$$

To find the change in the internal energy we have to find the mean value of the change of each of the energy levels of the system. According to the rules for averaging we have

$$\delta E = \overline{\delta \epsilon} = \sum \delta \epsilon_i w_i = - \sum f_i w_i \delta \lambda = - \Lambda \delta \lambda, \quad (22.3)$$

where Λ stands for the mean force acting overall on the system as the parameter λ changes, $\Lambda = (\sum f_i w_i)_{w_i}$.

The quantity $(-\Lambda d\lambda)$ represents the work done on the system as the parameter λ changes by $d\lambda$. The minus sign shows that the work is done by external forces on the system.

Let δW denote the mean value of the work done on the system as external parameters λ change. Then we have

$$(\delta E)_{w_i} = \delta W. \quad (22.4)$$

Consider, in particular, the important case where the linear dimension of the system, determined by the coordinate x , serves as the generalized coordinate. In this case instead of the generalized force it is convenient to introduce the pressure p , which we shall define as the mean force acting on 1 cm² normal to the surface of the body (the system), i.e.

$$p = \Lambda/S.$$

We then have

$$(\delta E)_{w_i} = \delta W = -pS\delta x = -p\delta V, \quad (22.5)$$

where δV is the change in the volume of the system.

Such a definition of the pressure is not new; we have made use of it in the kinetic theory of gases. In §8 we have defined the pressure as the mean force acting on unit surface of the wall due to gas molecules impinging on it. In a system containing a large number of particles the true force always has a value very close to its mean value. It is this that justifies the introduction of the pressure which, to a high degree of accuracy, can be substituted for the actual force acting on the surface of the body.

It is obvious that $(\delta E)_{w_i}$ does not represent the total possible change in the energy of the system, and is not the total differential of any expression. Indeed, for a given structure of the system the generalized force $\Lambda = \sum f_i w_i$ represents a function of the external parameters λ and the temperature θ . Hence we can write in more detail as follows:

$$(\delta E)_{w_i} = -\Lambda(\lambda, \theta) \delta \lambda. \quad (22.6)$$

The change in the energy as the parameter λ changes in the range from λ_1 to λ_2 , or the work done on the system, is equal in this case to

$$W = - \int_{\lambda_1}^{\lambda_2} \Lambda(\lambda, \theta) \delta \lambda.$$

The meaning of the integral in the above formula obviously depends on the path of integration, i.e. on the character of the transition from λ_1 to λ_2 . In particular, in the case when $\lambda = V$,

$$W = - \int_{V_1}^{V_2} p(V, T) \delta V. \quad (22.7)$$

Since the pressure depends on the volume and temperature, the transition from a volume V_1 to a volume V_2 by a different path of integration, i.e. for a different character of the transition from V_1 to V_2 , leads to a different value for the work W .

§ 23. The change in the energy of a system in the general case of a quasi-static process

Now consider the change in the energy of a subsystem in the more general case when it interacts with bodies surrounding it (the medium), exchanging energy with them by direct contact.

In what follows we shall restrict ourselves to processes in which the state of statistical equilibrium in the system is not violated. Such processes for which the system can be considered to be always in a state of statistical equilibrium or, more precisely, during the course of which the system passes through a sequence of equilibrium states, will be called quasi-static or reversible processes. The question as to how much the state of the system can in fact change without violating the state of equilibrium, i.e. whether quasi-static transitions be realized in the system, will be discussed below.

Since the system is in a state of equilibrium during all the time for which the process occurs, the probability distribution is determined by the Gibbs equilibrium distribution.

The total change in the mean energy can be written

$$\delta E = \delta \left(\sum_i \epsilon_i w_i \right) = \left(\sum_i w_i \delta \epsilon_i \right)_{w_i} + \left(\sum_i \epsilon_i \delta w_i \right)_\lambda, \quad (23.1)$$

where w_i is the Gibbs distribution for a temperature equal to that of the reservoir. The latter, however, need not remain constant in the course of the process.

The first term of formula (23.1) expresses, as before, the work done on the system.

The second term represents that part of the change in the energy of the system through interaction with the medium which is not connected with the change of the external parameters. In other words, the second term of (23.1) is equal to the change in the mean energy of the system due to the direct energy transfer from the particles of the medium to the particles of the system, and which is not accompanied by a change of external fields or the mutual positions of the bodies. This part of the change in the energy will be called the amount of heat absorbed by the system, and will be denoted by δQ . We then have

$$\delta E = \delta W + \delta Q. \quad (23.2)$$

Formula (23.2) represents the energy conservation law for thermal processes (the first law of thermodynamics). It is in this form that the energy conservation law was first established after Joule's experiments.

Statistical physics allows one to reveal the molecular meaning of the quantities entering into (23.2) and also, at least for the simplest systems, makes it possible to calculate them theoretically.

In order to elucidate the molecular meaning of the amount of heat we shall consider an arbitrary unclosed system in which a quasi-static process takes place. For this quasi-static process we can write, making use of the definition (23.2),

$$\delta Q = \left(\sum_i \epsilon_i \delta w_i \right)_\lambda = \delta E - \sum_i w_i \delta \epsilon_i = \delta E - \frac{\sum \exp(-\epsilon_i/\theta) \Omega(\epsilon_i) \delta \epsilon_i}{Z}.$$

The second term can be transformed in the following way. We have the obvious identity *

$$\delta \left(\sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i) \right) = -\frac{1}{\theta} \sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i) \delta \epsilon_i + \sum \epsilon_i e^{-\epsilon_i/\theta} \Omega(\epsilon_i) \frac{\delta \theta}{\theta^2},$$

from which it follows that

$$\begin{aligned} \sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i) \delta \epsilon_i &= \\ &= -\theta \delta \left(\sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i) \right) + \frac{\delta \theta}{\theta} \sum \epsilon_i e^{-\epsilon_i/\theta} \Omega(\epsilon_i), \end{aligned} \quad (23.3)$$

whence, dividing (23.3) by Z , we find

$$\frac{\sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i) \delta \epsilon_i}{Z} = -\frac{\theta \delta Z}{Z} + \frac{\delta \theta}{\theta} \frac{\sum \epsilon_i e^{-\epsilon_i/\theta} \Omega(\epsilon_i)}{Z}. \quad (23.4)$$

The first term of the right-hand side of (23.4) can be written in the form

$$-\theta \frac{\delta Z}{Z} = -\theta \delta \ln Z.$$

In the second term the expression $Z^{-1} \sum \epsilon_i \exp(-\epsilon_i/\theta) \Omega(\epsilon_i)$ can be replaced by $\bar{\epsilon}$ or E . We then have

$$\delta Q = \delta E + \theta \delta \ln Z - E \frac{\delta \theta}{\theta} = \theta \delta \left(\frac{E}{\theta} + \ln Z \right). \quad (23.5)$$

Thus, we arrive at the following important conclusion.

If in a macroscopic system a process takes place, in the course of which the system remains at all times in equilibrium with the reservoir, then the change in its energy can be written in the form

$$\delta E = \delta W + \delta Q = -\Lambda d\lambda + \theta \delta \left(\frac{E}{\theta} + \ln Z \right). \quad (23.6)$$

* In the differentiation the variable quantities are ϵ_i and θ . The number of states corresponding to a given energy obviously remains constant, and is characteristic of a given system of particles.

Formula (23.6), which has a basic significance in what follows, represents a general expression for the change of energy in a quasi-static process.

As is seen from formula (23.6), the change in the energy is resolved into two parts: the work δW done on the system (or by the system), and the amount of heat δQ absorbed (or given up) by the system.

The work done is connected with a change in the values of the allowed energy levels, which, as we have seen in the preceding paragraph [formula (22.1)], is due to a change in the external parameters. If, in particular, the system consists of individual independent particles and one can speak of the energies of the individual particles, then the work done is connected with a change in the energy of the individual particles.

If the external parameters do not change (the work of external forces is equal to zero), then the energy levels of the system remain unchanged. In this case the energy supplied to the system from without contributes to a change in the probability distribution. States with a higher energy become more probable — the system is heated. If, for example, the system represents an ideal gas, then as energy is supplied the number of molecules having relatively large energies increases, whereas that of molecules having small energies decreases. In the case where the system gives up energy, a reverse redistribution of probabilities takes place: states with a lower energy become more probable — the system is cooled.

We shall now discuss the question as to when a process can be considered quasi-static.

If the external conditions in which the system is found are changed, for example if its volume or the fields acting on it are changed, or if it obtains a certain amount of energy from outside by means of direct contact, then the state of equilibrium in the system will be violated. If then the system is isolated, then in the course of time the system will necessarily come into a state of statistical equilibrium. Indeed, we have said that the state of a complex system does not depend on its initial state. Hence, if the time of observation is sufficiently large, then the system spends most of the time in a state of statistical equilibrium, independently of the state in which it was initially. After the lapse of a relaxation time τ the system, which was initially in a non-equilibrium, highly improbable state, passes over into a more probable equilibrium state. The question as to how this transition will proceed and how much time it will take cannot, in general, be answered. Processes going on in the system in this case depend on the nature of the system and on the character of its equilibrium state.

We shall now assume that the change of the external conditions proceeds sufficiently slowly. That is to say, we shall assume that an appreciable

change of external conditions takes place for time intervals which are very large in comparison with the relaxation time. Then at every instant the system will manage to come into a state of equilibrium corresponding to the given external conditions.

We shall illustrate this by a simple example. Consider a process of compression and expansion of a gas under a piston. As the piston moves it does work on the parts of the gas contiguous to it. The corresponding molecules obtain an excessive energy in comparison with the remaining bulk of gas molecules, and the gas becomes non-uniform. Owing to collisions between molecules the non-uniformity will tend to vanish, and the energy supplied will be distributed uniformly between all the molecules of the gas. In order that this process may take place, and that the gas which has been disturbed from an equilibrium state may come back to it, a certain time, which is the characteristic relaxation time for a given process, is needed. If the piston is displaced so slowly (for example, by very weak and rare pushes) that the time needed for the displacement of the piston over an appreciable distance is very large in comparison with the relaxation time, then all perturbations of the uniformity of the gas will be resolved. The gas will at all times be uniform, i.e. will be in a state of equilibrium.

Analogously, in the case of heating, in the region adjacent to the source of heat (for example, adjacent to one of the walls of the container), a change in the velocity distribution of molecules will take place and the percentage of molecules with large velocities will increase. The non-uniformities of the gas will be smoothed out by collisions in a certain relaxation time. If the heating of the gas proceeds so slowly that an appreciable change in the temperature takes place over a time considerably larger than the relaxation time, then the non-uniformities will be resolved and the gas will at all times be in a state of equilibrium.

Thus, the condition of the quasi-static character of a process is the condition of its slowness. To every relaxation time there corresponds a proper rate of change of external conditions for which the process can be assumed to be quasi-static.

It goes without saying that the quasi-static process represents a certain idealization of real processes which always proceed with a finite velocity.

Every quasi-static process is a reversible process. This means that, if in the course of the process the system passed through a given sequence of equilibrium states (direct process), then it can also be brought into its initial state by passing through the same sequence of states in reverse (reverse process). For this the external conditions in which the system is found need only be changed in the reverse order.

It is impossible to do this for non-quasi-static processes. In a non-quasi-static process the equilibrium state of the system is violated. The state of a non-equilibrium system is not determined by giving the external parameters and temperature of the system, but needs the specification of a number of other quantities, for example, the temperature distribution or the density distribution inside the system. The change of the external conditions in the reverse sequence will not mean that the system passes through the same states in the reverse order. Hence non-quasi-static processes are irreversible.

It further goes without saying that a completely reversible process is an idealization. Real processes always proceed with a finite velocity and are accompanied by a violation of the equilibrium of the system. However, it is often possible, to a good enough approximation, to disregard small perturbations of the equilibrium state of the system and to consider the process, which is actually proceeding with a finite velocity, as a reversible process.

§ 24. Entropy and the basic thermodynamic equality

Formula (23.5) shows that for a quasi-static process the amount of heat absorbed or released by the system can be written in the form

$$\delta Q = \theta \delta \sigma, \quad (24.1)$$

where $\delta \sigma$ is the change in a certain function

$$\delta \sigma = \delta \left(\frac{E}{\theta} + \ln Z \right). \quad (24.2)$$

It is obvious that $\delta \sigma$ represents the total differential of the expression in the parentheses

$$\sigma = \frac{E}{\theta} + \ln Z + \text{const}, \quad (24.3)$$

where the constant is arbitrary. The function σ is called the entropy of the system. The physical meaning of this very important quantity will be explained somewhat later.

By means of the entropy the change in the energy of a system for a quasi-static process can be written in the form

$$\delta E = \theta \delta \sigma - \Lambda \delta \lambda. \quad (24.4)$$

The external parameter is most frequently the volume V of the system. Then

$$\delta E = \theta \delta \sigma - p \delta V. \quad (24.5)$$

Formula (24.4), or (24.5), expressing the change in the energy of a system in the most general case of a quasi-static process, is called the basic thermodynamic equality.

We have obtained the basic thermodynamic equality in a purely statistical way. However, this equality, as well as the entropy defined by formula (24.1), were introduced historically in phenomenological thermodynamics (see below).

The basic thermodynamic equality shows that the total change in the energy of a system in a quasi-static process is determined by the change in the external parameter $\delta \lambda$ and the change in the entropy $\delta \sigma$.

Indeed, from formulae (24.4) and (24.5) we find

$$\Lambda = - \left(\frac{\partial E}{\partial \lambda} \right)_{\sigma}, \quad p = - \left(\frac{\partial E}{\partial V} \right)_{\sigma}, \quad \theta = \left(\frac{\partial E}{\partial \sigma} \right)_{\lambda}, \quad (24.6)$$

so that we can write

$$\delta E = \theta \delta \sigma - \Lambda \delta \lambda = \left(\frac{\partial E}{\partial \sigma} \right)_{\lambda} \delta \sigma + \left(\frac{\partial E}{\partial \lambda} \right)_{\sigma} \delta \lambda. \quad (24.7)$$

Thus, the thermodynamic internal energy of a system can be considered as a function of the independent variables σ and λ (or V). As is seen from the equality $\theta = (\partial E / \partial \sigma)_{\lambda}$ and the condition $\theta > 0$, the energy is a monotonic function of the entropy.

The formula for the change in the entropy is similar in its structure to the formula connecting the change in the potential energy with the generalized coordinate in mechanics. For this reason the internal energy is called the thermodynamic potential with respect to the generalized coordinates σ and λ . The quantities θ and Λ play the role of generalized forces.

The quantity δE is a total differential, in contrast to the heat δQ and work δW , which in the general case do not represent the total differentials of any expressions.

Taking the integral over the closed cycle of changes in the state of the system, we find

$$\oint \delta E = 0.$$

This is quite natural, since the internal energy E is a single-valued state function of the system. When the system comes back into its initial state its energy will also assume its initial value.

The work W and the amount of heat Q depend not only on the state but also on the character of the processes taking place in the system. Hence it makes no sense to speak of the amount of heat in the system in a given state. Only the change δQ in the heat has a meaning.

Formula (24.1) shows that the ratio $(\delta Q)/\theta$ is a total differential and, consequently, the entropy σ represents a single-valued state function of the system. For it the condition

$$\oint \delta \sigma = \oint \frac{\delta Q}{\theta} = 0$$

holds.

The temperature θ can be considered from the mathematical point of view as the integrating divisor of the expression δQ .

In order to explain the physical meaning of the entropy, which will be done in the next paragraph, we have also to consider its other properties.

According to formula (24.3), to calculate the entropy it is necessary to know only the partition function Z . In formula (24.3) an integration constant appears. Thus, the entropy is determined only to within an arbitrary constant. It is important, however, as will be shown in §36, that this constant is indeed a constant, depending neither on the temperature of the system nor on any other parameter characterizing the state of the system (the volume, physical and chemical state of the system and so on). This constant can be chosen as zero, and formula (24.3) can be written in the form

$$\sigma = \frac{E}{\theta} + \ln Z. \quad (24.8)$$

We transform the above expression for the entropy, making use of the fact that in a state of statistical equilibrium the system will possess energies ϵ_i close to the mean energy $\bar{\epsilon}$ for an overwhelmingly large part of the time.

We transform the partition function of the system, taking into account that only the most probable state with the energy $\bar{\epsilon}$ gives a significant contribution to it. We can write

$$Z = \sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i) \approx e^{-\bar{\epsilon}/\theta} \Omega(\bar{\epsilon}) = e^{-E/\theta} \Omega(\bar{\epsilon}). \quad (24.9)$$

Here we have retained only the largest term in the sum over the states.

Substituting (24.9) into (24.8), we find

$$\sigma = \ln \Omega(\bar{\epsilon}). \quad (24.10)$$

The entropy of a macroscopic quasi-closed system turns out to be equal to the logarithm of the number of states corresponding to the mean energy of the system, i.e. to the logarithm of the number of states of the system in a state of statistical equilibrium.

Thus, the entropy σ is identical with the function σ which we have introduced in § 16 (formula (16.7)). A very important property of the entropy is its additivity. The entropy of a complex system in an equilibrium state, consisting of n subsystems, is equal to

$$\sigma = \ln \Omega = \ln \prod_n \Omega_n = \sum \ln \Omega_n = \sum \sigma_n. \quad (24.11)$$

The additivity of the entropy follows directly from its definition (16.7) or (24.3).

The degree of accuracy of the statement concerning the additivity of the entropy is the same as that of the statement concerning the additivity of the energy.

§ 25. The law of increase of entropy

In the preceding paragraphs we have considered a quasi-closed system in a state of statistical equilibrium of performing a quasi-static (reversible) process. We have established that the molecular (statistical) interpretation can be given to a number of macroscopic notions: the internal energy, work, amount of heat. The macroscopic quantity entropy also has a molecular interpretation.

Formulae (24.8) and (24.10) allow one to calculate the value of the entropy, but do not shed any light on the meaning of this quantity. In order to elucidate the molecular meaning of entropy one has to consider a system with a simpler statistical behaviour than a quasi-closed system, i.e. a closed system. The simplicity of a closed system will allow us not only to study the

properties of equilibrium systems, but also to include the treatment of non-equilibrium systems.

Imagine a closed macroscopic system as a set of a large number of parts. Each of the parts has dimensions small in comparison with those of the system as a whole, but still contains a large number of particles and is a quasi-closed system. Since our division is quite arbitrary, it can always be made. Assume that all the parts of our complex system have come into a state of statistical equilibrium. Then for each of these one can write the expression (24.10) for the entropy:

$$\sigma_n = \ln \Omega_n(\bar{\epsilon}_n), \quad (25.1)$$

where the index n denotes the number of the part.

However, we shall not assume that there is statistical equilibrium between the parts of the system. For example, different parts of the system may have different temperatures although the proper temperature of each of the parts is constant. The entire closed system as a whole will be in a non-equilibrium state.

We shall determine the entropy of the closed non-equilibrium system.

According to the very meaning of the concept the entropy of a complex system should be assumed to be made up additively of the entropies of all parts constituting it, i.e.

$$\sigma = \sum_n \sigma_n. \quad (25.2)$$

As we have seen above, this formula is trivial for the case of a system in which there is equilibrium between the parts. It represents a natural generalization of the notion of entropy to the case of a non-equilibrium system.

For each of the parts of the system the equilibrium value of the entropy can be written from formula (24.10). We then have

$$\sigma = \sum \sigma_n = \sum \ln \Omega_n(\bar{\epsilon}) = \ln \prod_n \Omega_n(\bar{\epsilon}_n) = \ln \Omega, \quad (25.3)$$

where $\Omega = \prod_{n=1}^N \Omega_n(\bar{\epsilon}_n)$ is the total number of states of a system consisting of N independent parts.

We see that the entropy of a closed system turns out to be equal to the logarithm of the number of states of the system. This may not be the same as the logarithm of the number of states $\Omega(\bar{\epsilon})$ of the entire system when it is in

a state of statistical equilibrium (which always holds for each of its parts or for the entire system in a state of statistical equilibrium).

In the closed system considered here the microcanonical distribution (16.1) holds, connecting the probability of a state of the closed system with the number of its states Ω . Expressing Ω in terms of w , we find

$$\sigma = \ln w + \text{const.} \quad (25.4)$$

Formula (25.4), representing the basis of the statistical treatment of thermodynamics, is called the Boltzmann formula.

The Boltzmann formula connects the value of the entropy for a given state of a closed system with the probability of that state. The change in the entropy as the closed system passes from one state into another is equal to

$$\sigma_2 - \sigma_1 = \Delta\sigma = \ln \frac{w_2}{w_1}, \quad (25.5)$$

where w_1 and w_2 , σ_1 and σ_2 are the probabilities and entropies of the first and second state, respectively.

The entropy of a quasi-closed system can be expressed in terms of the probability density $\rho(p, q)$ entering into the classical Gibbs distribution (19.4) in the following way.

The function ρ satisfies the normalization condition

$$\int \rho d\Omega = 1.$$

Taking into account that $\rho(\epsilon)$ has a sharp maximum at $\epsilon = \bar{\epsilon}$, we have approximately:

$$\int \rho d\Omega \approx \rho(\bar{\epsilon}) \Omega(\bar{\epsilon}) = Z^{-1} e^{-\bar{\epsilon}/\theta} \Omega(\bar{\epsilon}) = 1.$$

Hence the entropy of the system, on the basis of (24.10), is equal to

$$\sigma = \ln \Omega(\bar{\epsilon}) = \ln \frac{1}{\rho(\bar{\epsilon})} = -\ln \rho(\bar{\epsilon}) = \ln Z + \frac{\bar{\epsilon}}{\theta}.$$

On the other hand, one can write

$$\overline{\ln \frac{1}{\rho(\epsilon)}} = \int \rho(\epsilon) \ln \frac{1}{\rho(\epsilon)} d\Omega = \int \rho(\epsilon) \left(\ln Z + \frac{\epsilon}{\theta} \right) d\Omega = \ln Z + \frac{\bar{\epsilon}}{\theta}.$$

Hence the entropy σ of a quasi-closed system can be written in the form

$$\sigma = \ln \frac{1}{\rho(\epsilon)}. \quad (25.6)$$

We shall need this formula in what follows.

We shall return to the Boltzmann formula (25.4) and see how, by means of it, the laws of the change of state of a closed system in time can be established.

Assume that the closed system was initially in a certain non-equilibrium state. Then w_1 stands for the probability of the initial non-equilibrium state. After the lapse of the relaxation time the system will pass over from the non-equilibrium state into a state of statistical equilibrium. This transition takes place due to a weak, but ever present, interaction between its parts. Without going into the question of how and in what time the equilibrium is established (this is a problem of physical kinetics), we can state that this transition occurs inevitably in any macroscopic system after the lapse of the relaxation time. By definition, the probability w_2 of the state of statistical equilibrium (in which the macroscopic system spends almost all its time) has the maximum value, so that $w_2 > w_1$. From formula (25.5) it follows that the entropy of a closed system increases as the latter passes from a non-equilibrium state to an equilibrium state. Thus, the increase in the entropy of a closed system turns out to be connected with its transition from a less probable state into a more probable one. The entropy of a system in a state of total statistical equilibrium has the highest value. The result obtained can be formulated in the following way.

If a closed macroscopic system is initially in a non-equilibrium state, the probability of this state and hence its entropy do not have the highest possible values.

The behaviour of the system which is most probable is that in which it will pass, after the lapse of the relaxation time, into its most probable state, whose entropy is a maximum.

It can be shown that, on the average, this transition will be performed monotonically, i.e. that the system comes into a state of statistical equilibrium by passing successively through a number of more and more probable states until it reaches the state of complete equilibrium. The entropy of the system increases progressively, attaining its maximum value in the most probable equilibrium state. Thus, the change of entropy in time proceeds as is shown in fig. III.12 by the solid (not the dotted!) curve.

Now imagine a case where a closed system is initially already in the state

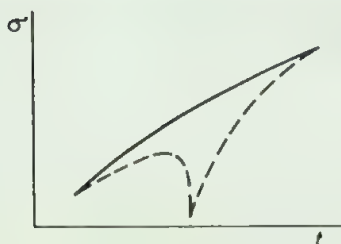


Fig. III.12

of total statistical equilibrium, in which its entropy has the maximum value. Then in the course of a very long time, exceeding the relaxation time, the system will remain in its equilibrium state, and the entropy will preserve its maximum value. In general, it can be said that the most probable trend of processes in a closed macroscopic system is that in which the entropy increases or remains constant:

$$\Delta\sigma \geq 0, \quad (25.7)$$

where the inequality sign refers to processes bringing the system nearer to a state of statistical equilibrium, and the equality sign refers to processes taking place in a system which is already in a state of equilibrium.

We know, however, that probability predictions, as applied to macroscopic systems, are practically speaking of a completely reliable character. Hence, leaving the problem of a more complete and deeper treatment of the law of increase of entropy until §36, we shall take into account only the most probable trend of the entropy and assume that formula (25.7) does not have simply a likely character, but a completely reliable one.

Then the change or the constancy of the entropy can be considered as a criterion of irreversibility or reversibility of the processes taking place in a closed system. In irreversible processes, in the course of which the system approaches an equilibrium state, the entropy increases, whereas in reversible processes it remains constant.

As an important example of an irreversible process taking place in a closed system, we shall consider a process occurring when parts of the system having different temperatures are brought into contact.

If two parts of the system, having temperatures θ_1 and θ_2 (for concreteness we assume that $\theta_2 > \theta_1$) are in contact, then the change in the entropy of the closed system is equal to

$$\delta\sigma = \delta\sigma_1 + \delta\sigma_2 = \frac{\partial\sigma_1}{\partial E_1} \delta E_1 + \frac{\partial\sigma_2}{\partial E_2} \delta E_2 = \frac{\delta E_1}{\theta_1} + \frac{\delta E_2}{\theta_2} \geq 0.$$

Since the system is closed, its total energy is conserved, so that

$$\delta E = \delta E_1 + \delta E_2 = 0.$$

Consequently,

$$\delta\sigma = \delta E_1 \left(\frac{1}{\theta_1} - \frac{1}{\theta_2} \right) \geq 0. \quad (25.8)$$

Formula (25.8) shows that, if $\theta_2 > \theta_1$, then it follows from the law of increase of entropy that $\delta E_1 \geq 0$. This means that the first part, having a lower temperature, absorbs energy from the second part. In other words, heat always passes from a hotter body to a colder one.

§26. The basic thermodynamic inequality

It is natural to generalize the law of increase of entropy to the case of open systems.

Such a generalization can easily be made in the case of system which are not closed but are thermally isolated. We shall understand thermally isolated systems to be those for which the interaction with surrounding bodies reduces to an action due to external fields, i.e. to a change of the external parameters.

A change in the external fields, as was explained in §22, can lead to a change in the energy levels of the system (or to a change in the energy of individual particles in the case of gases), but it does not lead to any change in the probability distribution. Hence the transitions from less probable states to more probable ones in a thermally isolated system obey the same laws as in a closed system. The results of the preceding section can be applied directly to the case of systems which are thermally isolated but not closed, by writing the law of increase of entropy for these as

$$\delta\sigma \geq 0. \quad (26.1)$$

In the general case of open systems exchanging energy in an arbitrary way with bodies surrounding them the following inequality can be written:

$$\delta\sigma \geq \frac{\delta Q}{\theta}. \quad (26.2)$$

For quasi-static processes this goes over into the equality (24.1), while in the transition to a thermally isolated system it goes over into the inequality (26.1).

The inequality (26.2) means physically that in irreversible processes the entropy of the system increases by an amount larger than $(\delta Q)/\theta$ which is the amount by which the entropy increases because the system absorbs heat. This excess of the entropy increase in comparison with $(\delta Q)/\theta$ is associated with the transition to a more probable state, i.e. with the approach to equilibrium.

Combining (26.2) with the basic thermodynamic equality, one can write the following basic thermodynamic inequality for the general case of arbitrary processes in open systems:

$$\delta E \leq \theta \delta\sigma + \delta W, \quad (26.3)$$

where the equality sign refers to reversible processes, while the inequality sign refers to irreversible processes.

The basic thermodynamic inequality unifies the concepts of the energy conservation law and the entropy increase law, and can be called the unified form of the expression of the first and second laws of thermodynamics.

The relations obtained allow us to point the way to determine the scale of the statistical and absolute temperature.

The statistical temperature is measured in ergs, whereas in practice, for the measurement of the temperature, use is made of another system of units: degrees.

Of great importance is the absolute scale of temperature, in which the temperature is measured from an absolute zero and which is, essentially, identical with the statistical temperature.

To establish the relationship between the statistical and absolute temperatures it is only necessary to find the numerical expression of the units of energy in terms of degrees. Thus, one can write

$$\theta = kT, \quad (26.4)$$

where the constant k represents the conversion factor relating ergs to degrees. It appears to be a universal constant the numerical value of which can be obtained only from experiment.

The quantity k is called Boltzmann's constant. Measurements (for example, of the thermal capacity of gases) show that $k = 1.38 \times 10^{-16}$ erg deg $^{-1}$.

Making use of the absolute scale of temperatures and introducing the entropy expressed in erg deg $^{-1}$, $S = k\sigma$, we can rewrite formulae (25.4), (24.3) and (24.5) in the form

$$S = k \ln w + \text{const} = \frac{E}{T} + k \ln Z + \text{const}, \quad (26.5)$$

$$\delta E = T\delta S - p\delta V. \quad (26.6)$$

§ 27. The maximum work to be obtained from thermal processes. The impossibility of constructing a perpetual motion machine of the second kind. Phenomenological definition of the entropy

We can now consider problems whose study was, historically, a stimulus to the development of phenomenological thermodynamics. The objective is to calculate the amount of useful work which can be obtained in changing the internal energy of a system. In thermodynamics we call devices for obtaining work, heat engines.

All heat engines can be divided into two types. Engines of the first type produce useful work by a sequence of closed cycles. These include steam-engines, steam turbines and gas turbines, compressors, internal-combustion engines etc. At the end of each cycle the engine comes back to the initial state.

Hence the engine itself serves as a kind of transmission mechanism conducive to the transformation of the internal energy of the work substance into work.

Engines of the second type perform non-cyclic processes and produce useful work. Such an engine — a system which is initially in a non-equilibrium state — comes into an equilibrium state. The transition to the equilibrium state yields useful work.

All engines of a once-only action belong to this type. In such engines the useful work is most often obtained from chemical reactions going on in the system. As an example we can cite galvanic cells, rockets etc.

We shall first consider heat engines performing closed cycles (those of the first type).

Within the framework of this book we cannot, of course, study in detail the theory of the action of particular heat engines. This is a problem of technical thermodynamics. We shall confine ourselves only to the explanation of the principal aspect of the problem.

That is, we shall consider, first of all, the question as to whether the internal energy — the energy of thermal motion of the particles constituting the body — can be transformed directly into useful work.

We shall show that the existence of such an engine, which we have called above a perpetual-motion machine of the second kind, contradicts the law of increase of entropy and is hence impossible. For this we shall consider an arbitrary thermally isolated system having an initial energy E_0 , entropy S_0 and external parameters λ_0 . We assume that the system, remaining thermally isolated, passes over in a non-quasi-static way on account of a change in the external parameters into a new state with an energy E' , entropy S' and parameters λ' . Thereafter the system returns in a quasi-static way to the state with external parameters λ_0 . However, in the final state it will have entropy S' and energy E' differing from the initial values of the entropy and energy.

According to the law of increase of entropy, $S' \geq S_0$, for a thermally isolated system. But from the condition $(\partial E / \partial S)_\lambda = T > 0$, expressing the monotonic behaviour of the energy as a function of the entropy, it follows that an increase in the internal energy of the body, i.e. $E' \geq E_0$, corresponds to an increase in the entropy.

In the course of the process considered the energy of the thermally isolated body must increase. An increase in the energy can take place only at the expense of work done on the system by external bodies.

Thus, from the law of increase of entropy it follows that the system considered not only cannot serve as a source of useful work but, that on the contrary, work must be done on the system when it goes over irreversibly into a state with a new energy.

Hence we can state that the law of increase of entropy is equivalent to the proposition of the impossibility of creating a perpetual-motion machine of the second kind.

Of course, the reverse statement is also valid: from the impossibility of creating a perpetual-motion machine of the second kind there follows unambiguously the existence of a monotonically increasing state function, the entropy, for a closed system (see below).

Proceeding from this principle, one could introduce into thermodynamics the entropy and the law of its increase as a quantitative expression of the

second law of thermodynamics. It is in just this way that thermodynamics developed, and this historical order of developing thermodynamics is also preserved in contemporary text-books. Thus, the historical development of thermodynamics was in the reverse of the order in which we develop the material in this book.

Returning to the consideration of the problem of obtaining work, we note that to obtain useful work it is necessary to have at least two bodies at different temperatures T_1 and T_2 , i.e. a system of bodies which are not in equilibrium.

Before calculating the work obtained, we shall show that the maximum work is obtained in a reversible (quasi-static) process.

Let the change in the energy in the general case be equal to

$$\delta E = \delta Q + \delta W.$$

For a reversible process the same change in the energy can be written in the form

$$\delta E = T\delta S + \delta W'.$$

Subtracting, we find

$$\delta W' - \delta W = T\delta S - \delta Q.$$

But $T\delta S > \delta Q$, so that

$$\delta W' - \delta W > 0,$$

or

$$\delta W' > \delta W. \quad (27.1)$$

The maximum work is obtained in a reversible (quasi-static) transition. To calculate this work we first note that the establishment of thermal contact between bodies at different temperatures leads to an irreversible heat transfer and to no useful work.

Hence the working engine must contain three elements:

- (1) a system with a temperature T_2 (a hot reservoir),
- (2) an auxiliary system by means of which the energy is transferred from a warmer body to colder one without direct contact between them (a working substance),
- (3) a system with a temperature $T_1 < T_2$ (a cold reservoir).

An energy $\delta E_2 = \delta Q_2$ is transferred from the hot reservoir to the working substance in a reversible way. For this it is necessary that the temperature of the hot reservoir should be equal to that of the working substance during the course of the entire process of heat transfer (isothermal process). Then $\delta Q_2 = T_2 \delta S_2$.

A part of the energy obtained from the hot reservoir must be transferred to the cold reservoir (otherwise we would obtain a perpetual-motion machine of the second kind), while another part of the energy is transformed into useful work. The energy balance reads:

$$\delta Q_2 + \delta Q_1 = -\delta W.$$

In order that irreversible processes may be avoided the heat δQ_1 must be transferred to the cold reservoir in an isothermal way at the temperature T_1 of the cold reservoir. Hence the working substance must pass over from a temperature T_2 to a temperature T_1 in a thermally isolated and reversible way, and thereupon it must transfer to the cold reservoir an amount of heat δQ_1 in a quasi-static way. In order that the process may be repeated the working substance must return in a thermally isolated way (adiabatically) to the temperature T_2 . This closed cycle is called the Carnot cycle.

Since all processes in the system (hot reservoir+working substance+cold reservoir) are reversible, the total change in the entropy is

$$\delta S = \delta S_1 + \delta S_2 = 0.$$

For the work done one can write that

$$\begin{aligned} -\delta W &= \delta Q_2 + \delta Q_1 = T_2 \delta S_2 + T_1 \delta S_1 \\ &= (T_2 - T_1) \delta S_2 = \frac{T_2 - T_1}{T_2} \delta Q_2 = \frac{T_2 - T_1}{T_2} \delta E_2. \end{aligned}$$

The ratio of the work done to the amount of energy absorbed from the hot reservoir is called the efficiency η . In our case

$$\eta = \eta_{\max} = \frac{-\delta W}{\delta E_2} = \frac{\delta Q_2 + \delta Q_1}{\delta Q_2} = \frac{T_2 - T_1}{T_2}. \quad (27.2)$$

It is clear from the nature of the above derivation that the efficiency obtained has the maximum possible value. If irreversible processes take place in the heat engine then we will always have $\eta < \eta_{\max}$.

Thus, the maximum efficiency is possessed by a reversible engine working according to the closed Carnot cycle. The value of the efficiency does not depend on the nature of the working substance and is determined solely by the ratio of the temperature drop $T_2 - T_1$ to the temperature of the hot reservoir T_2 .

We shall now reproduce briefly the train of reasoning which led to the introduction of the notion of entropy into phenomenological thermodynamics. It was, to a certain degree, the reverse of ours.

From the relation (27.2) it follows that

$$\frac{\delta Q_2}{T_2} + \frac{\delta Q_1}{T_1} = 0. \quad (27.3)$$

The ratio of the amount of heat δQ obtained at a certain temperature T to the value of the temperature, $T^{-1}\delta Q$, was called by Clausius the reduced heat. Consequently, the algebraic sum of reduced heats for the Carnot cycle is equal to zero. We have obtained this result by means of formulae relating a change in the amount of heat to the absolute temperature and entropy. However, for an ideal gas the amount of heat absorbed or released in an isothermal process and the change in the temperature in an adiabatic process can be found directly. This allows one to find the efficiency of a reversible engine operating according to the Carnot cycle with an ideal gas as the working substance, and the result agrees, of course, with (27.2). Thus, formulae (27.2) and (27.3) can also be obtained without introducing the entropy.

Now let us consider a heat engine performing an arbitrary reversible cycle. This cycle can be divided into an infinitely large number of infinitely small Carnot cycles. Summing the relation (27.3) over all the elementary cycles, we can write

$$\oint \frac{dQ}{T} = 0. \quad (27.4)$$

Hence it follows that the quantity $T^{-1}dQ$ represents the total differential of a certain state function S of the system. For a cyclic process the total change in the function S is equal to zero. The function S is called the entropy.

The law of the constancy of the entropy for a reversible process in a closed system (which does not absorb or release any heat) follows directly

from the definition of entropy. To find the change in the entropy in an irreversible process, the transition from a certain initial state A into a final state B is considered for two modes, a reversible and an irreversible one. The change in the internal energy, which is also a state function, is equal to

$$\delta E = E_A - E_B$$

and does not depend on the mode of the transition.

The change in the entropy for a reversible mode is connected with the heat δQ absorbed, by the relation $\delta S = T^{-1}\delta Q$. Hence for a reversible change

$$\delta E = T\delta S + \delta W_{\text{rev}} = T\delta S - |\delta W_{\text{rev}}|, \quad (27.5)$$

where $|\delta W_{\text{rev}}|$ is the work done by the system on external bodies. For a transition in an irreversible mode the work produced will be smaller than in the case of a reversible mode (otherwise the efficiency of the irreversible closed cycle would be greater than that of the Carnot cycle). Hence, taking into account that

$$\delta E = \delta Q + \delta W_{\text{irrev}} = \delta Q - |\delta W_{\text{irrev}}|, \quad (27.6)$$

and subtracting (27.5) from (27.6), we find

$$T\delta S = \delta Q + |\delta W_{\text{rev}}| - |\delta W_{\text{irrev}}|,$$

or, since $|\delta W_{\text{rev}}| > |\delta W_{\text{irrev}}|$,

$$T\delta S > \delta Q. \quad (27.7)$$

Hence it follows that the change in the entropy for a transition $A \rightarrow B$ in an irreversible mode is

$$\delta S > \frac{\delta Q}{T}. \quad (27.8)$$

In a closed system an irreversible transition is accompanied by an increase in the entropy:

$$\delta S > 0. \quad (27.9)$$

Thus, proceeding from the fact of the impossibility of creating a perpetual-motion machine of the second kind, we arrive at the condition (27.3). The law of increase of entropy is obtained from (27.3) as a direct consequence, and is hence equivalent to the initial premise.

§28. The maximum work in non-cyclic processes and the thermodynamic potentials

We shall now consider the problem of the maximum work which can be done by a system performing a non-cyclic process (a heat engine of the second kind). Let a certain system (which we shall call the basic system) be in a hot reservoir in which a constant temperature T_0 and pressure p_0 are maintained. There is an interaction between the system and the reservoir, an exchange of heat and work. In addition to the basic system and the reservoir let there also be a thermally isolated body on which the system can do mechanical work. We shall call this body the object of the work, and the work done on it the useful work.

Let the basic system pass from an initial state into a certain final state, doing useful work $(-\delta W)$.

If the system did not interact with the reservoir the useful work $(-\delta W)$ would be equal to the change in its energy δE . However, the continuous interaction of the basic system with the reservoir as it does the useful work changes this relationship in an important way. That is, while the basic system does useful work the reservoir can in its turn exchange energy with the system.

Therefore the energy balance in the closed system (basic system+reservoir+object of the work) should be written in the form

$$\delta E + \delta E_0 = \delta W, \quad (28.1)$$

where δE_0 is the change in the energy of the reservoir, which can be written as

$$\delta E_0 = \delta Q_0 + \delta W_0.$$

Here δQ_0 is the heat transferred by the reservoir to the basic system, and δW_0 is the work done by the reservoir on the system. The size of the reservoir is so large that for any interaction with the system an infinitely slow quasi-static process takes place in it. The reservoir is in an equilibrium state

with temperature T_0 and pressure p_0 , which is not violated for any process in the basic system.

Hence for the reservoir one can write that

$$\delta E_0 = T_0 \delta S_0 - p_0 \delta V_0. \quad (28.2)$$

Since the volume of a closed system (a basic system+reservoir) must remain constant, we have

$$\delta V_0 + \delta V = 0. \quad (28.3)$$

From (28.1), (28.2) and (28.3) we find

$$-\delta W \geq \delta E - T_0 \delta S_0 + p_0 \delta V. \quad (28.4)$$

We write the law of increase of entropy in the closed system (basic system+reservoir) in the form

$$\delta S + \delta S_0 \geq 0.$$

Replacing δS_0 in (28.4) by δS , we find

$$-\delta W \geq \delta E - T_0 \delta S + p_0 \delta V = \delta R, \quad (28.5)$$

where the quantity R is equal to

$$R = E + p_0 V - T_0 S. \quad (28.6)$$

From what was said in the preceding paragraph, the maximum amount of useful work can be done on the object of the work in a reversible process, in the given case a reversible process in the system, since any process in the reservoir is always a reversible one. In this case in (28.4) only the equality sign should be retained, and we arrive at the relation

$$(\delta W)_{\max} = - \{ \delta E + p_0 \delta V - T_0 \delta S \} = - \delta R. \quad (28.7)$$

Thus, the maximum useful work is, in its absolute value, equal to the decrease in R . R involves quantities referring to the system (namely: E, V, S) as well as those referring to the reservoir (p_0, T_0).

A concrete expression for δW_{\max} containing only the characteristic parameters of the system can be obtained only for processes of a special type taking place in the system.

Assume that a system performs an isothermal process $T = T_0 = \text{const}$ and that the volume of the system does not change. In the case of a system in an external field of force at given T and V the state of the system is completely determined. If, however, the system is in an external field or is non-homogeneous, for example represents a mixture of reactants, then for given T and V the state of the system can change. Then the work obtained is

$$-\delta W \geq \delta(E - TS) = \delta F, \quad (28.8)$$

where

$$F = E - TS. \quad (28.9)$$

The quantity F , which is a measure of the work that can be obtained in an isothermal-isovolumic process taking place in a system interacting with a reservoir is called the free energy of the system. We see that only a part of the internal energy of the system can be used to obtain useful work. A part equal to TS and called the bound energy remains in the system.

Another important case is a process taking place at a constant temperature $T = T_0$ and a constant pressure $p = p_0$. In this case

$$-\delta W \geq \delta(E + pV - TS) = \delta G, \quad (28.10)$$

where

$$G = E + pV - TS, \quad (28.11)$$

is called the Gibbs thermodynamic potential.

The thermodynamic potential is a measure of the work done in an isothermal-isobaric process, just as the free energy is a measure of the work done in an isothermal-isovolumic process and the internal energy is a measure of the work in a thermally insulated system.

It can easily be shown that the expressions obtained are valid not only at constant temperature and pressure or volume but also in the case where the equalities $T = T_0$ and $p = p_0$ or $T = T_0$ and $\Delta V = 0$ hold only in the initial and final state of the system. Indeed, for example, for $T_{\text{init}} = T_{\text{fin}} = T_0$ and $V_{\text{init}} = V_{\text{fin}}$ we have

$$-\Delta W = (E - T_0 S + p_0 V_{\text{fin}}) - (E - T_0 S + p_0 V_{\text{init}}) = \Delta F.$$

§ 29. Properties of the thermodynamic potentials

Consider the case when the work $-\delta W$ done by a system in contact with a reservoir is equal to zero.

$$\delta R = \delta(E - T_0 S + p_0 V) \leq 0. \quad (29.1)$$

The equality sign refers to reversible processes, while the inequality sign refers to irreversible processes. The quantity R does not increase in any processes taking place in a system interacting with its reservoir.

For particular cases expression (29.1) is simplified.

For a closed system $\delta E = 0$ and $\delta V = 0$, so that (29.1) goes over into the previous relation:

$$\delta S \geq 0. \quad (29.2)$$

Other important cases are the isothermal-isovolumic process and the isothermal-isobaric process taking place in a system when the temperature or pressure of the system are equal to the corresponding quantities for the reservoir. In the first case $T = T_0$ and $\delta V = 0$, so that the inequality

$$\delta(E - T_0 S) = \delta F \leq 0. \quad (29.3)$$

holds. In the second case $T = T_0$ and $p = p_0$. Then

$$\delta(E - T_0 S + p_0 V) = \delta G \leq 0. \quad (29.4)$$

Thus, in an irreversible isothermal-isovolumic process taking place in a system interacting with the reservoirs its free energy decreases. In a reversible isothermal-isovolumic process the free energy remains constant. The free energy is an analogue of the entropy and, like the entropy, serves as a criterion of the reversibility or irreversibility of a process.

If, for example, a substance dissolves isothermally in a considerable volume of a solvent, then the temperature and volume of the system remain constant. The free energy of the solution will be smaller than that of the solvent and the substance dissolved, so that the process is irreversible.

Analogous properties are possessed by the thermodynamic potential for the isothermal-isobaric process. Isothermal-isobaric processes are in practice encountered rather frequently, because from the experimental point of view it is always easier to realize conditions for maintaining a constant pressure than those for maintaining a constant volume. For example, in the case of chemical reactions it is much simpler to preserve a constant pressure in the reaction vessel than to maintain a constant volume of the mixture of reagents.

The free energy and the thermodynamic potential play a fundamental role in thermodynamics. From the inequalities (29.3) and (29.4) it is seen that they replace the entropy in the case of open systems, while from (28.8) and (28.10) it follows that they are at the same time analogues of the internal energy.

Let us write the expressions for the changes in the free energy and the thermodynamic potential in a reversible process. In the general case formula (26.6) has the form

$$\delta E = T\delta S - p\delta V. \quad (29.5)$$

If we subtract from it $\delta(TS)$, then according to the definition of the free energy, we find

$$\delta F = -S\delta T - p\delta V. \quad (29.6)$$

Thus, the free energy is a function of the variables T and V (or λ). From formula (29.6) we obtain

$$S = -\left(\frac{\partial F}{\partial T}\right)_V, \quad (29.7)$$

$$p = -\left(\frac{\partial F}{\partial V}\right)_T. \quad (29.8)$$

These formulae are analogous to formulae (24.6).

Hence the free energy is a potential with respect to the variables T , V and λ . The quantities S , p and Δ , obtained from F by differentiation, play the role of generalized forces.

Formula (29.8) is particularly important. It determines the dependence of the pressure on the volume and temperature, i.e. it represents an equation of state.

Adding the total differential $\delta(pV)$ to (29.6) and taking into account the definition of the thermodynamic potential, we have

$$dG = \delta(E - TS + pV) = -S\delta T + V\delta p. \quad (29.9)$$

Thus, the Gibbs thermodynamic potential is a potential with respect to the variables T and p . The quantities S and V play the role of generalized forces:

$$S = -\left(\frac{\partial G}{\partial T}\right)_p, \quad (29.10)$$

$$V = \left(\frac{\partial G}{\partial p}\right)_T. \quad (29.11)$$

Since in practice it is most convenient to change or maintain temperature and pressure constant, the Gibbs thermodynamic potential is used especially often and is sometimes called the basic potential.

The quantity

$$H = E + pV \quad (29.12)$$

called the enthalpy is the potential with respect to the pair of variables p and S . From the enthalpy it is easy to obtain

$$\delta H = T\delta S + V\delta p. \quad (29.13)$$

Hence

$$T = \left(\frac{\partial H}{\partial S}\right)_p, \quad (29.14)$$

$$V = \left(\frac{\partial H}{\partial p}\right)_S. \quad (29.15)$$

If the state of the system depends on other external parameters λ , besides the volume, then formulae (29.6), (29.9) and (29.13) can be generalized and written in the form

$$\delta F = -S\delta T - p\delta V - \Lambda\delta\lambda, \quad (29.16)$$

$$\delta G = -S\delta T + V\delta p - \Lambda\delta\lambda, \quad (29.17)$$

$$\delta H = T\delta S + V\delta p - \Lambda\delta\lambda, \quad (29.18)$$

and, correspondingly,

$$\Lambda = -\left(\frac{\partial F}{\partial \lambda}\right)_{T,V} = -\left(\frac{\partial G}{\partial \lambda}\right)_{S,V} = -\left(\frac{\partial H}{\partial \lambda}\right)_{T,p}. \quad (29.19)$$

As will be seen from what follows, thermodynamic potentials and their derivatives completely determine the thermodynamic behaviour of an arbitrary system. We shall consider below the methods of determining the thermodynamic potentials theoretically and experimentally. However, it is necessary to obtain beforehand a number of thermodynamic relations connecting thermodynamic potentials and their derivatives with each other and with directly measured quantities.

§30. Some thermodynamic relations

The heat capacity at a constant volume C_V and the heat capacity at a constant pressure C_p , which are defined by the relations

$$C_V = \left(\frac{\partial E}{\partial T}\right)_V, \quad (30.1)$$

$$C_p = \left(\frac{\partial H}{\partial T}\right)_p \quad (30.2)$$

play a very important role in thermodynamics. Making use of formula (26.6), we find

$$C_V = \left(\frac{\partial Q}{\partial T}\right)_V = T\left(\frac{\partial S}{\partial T}\right)_V. \quad (30.3)$$

Analogously,

$$C_p = \left(\frac{\partial Q}{\partial T}\right)_p = T\left(\frac{\partial S}{\partial T}\right)_p. \quad (30.4)$$

From the definitions of the heat capacities it is clear that they are additive quantities. It is usually convenient to make use of molar heat capacities related to 1 gram-mole of a substance. In what follows, if not specified otherwise, we shall make use of molar heat capacities. The heat capacities represent thermodynamic characteristics of a substance which can be measured directly.

Another important relation is formula (29.8). Since the free energy is a function of the independent variables T and V , formula (29.8) can be written in the form

$$p = -\left(\frac{\partial F}{\partial V}\right)_T = f(T, V). \quad (30.5)$$

It determines the dependence of the pressure on the temperature and volume, i.e. it represents the equation of state of the body.

We now differentiate formulae (24.6), (29.7), (29.10) and (29.14) for the second time, forming mixed second derivatives. We have, obviously,

$$\left(\frac{\partial T}{\partial V}\right)_S = \frac{\partial}{\partial V} \left(\frac{\partial E}{\partial S}\right) = \frac{\partial}{\partial S} \left(\frac{\partial E}{\partial V}\right) = -\left(\frac{\partial p}{\partial S}\right)_V, \quad (30.6)$$

and, analogously,

$$\left(\frac{\partial S}{\partial V}\right)_T = -\frac{\partial^2 F}{\partial V \partial T} = \left(\frac{\partial p}{\partial T}\right)_V, \quad (30.7)$$

$$\left(\frac{\partial S}{\partial p}\right)_T = -\frac{\partial^2 G}{\partial p \partial T} = -\left(\frac{\partial V}{\partial T}\right)_p, \quad (30.8)$$

$$\left(\frac{\partial T}{\partial p}\right)_S = \frac{\partial^2 H}{\partial p \partial S} = \left(\frac{\partial V}{\partial S}\right)_p. \quad (30.9)$$

Formulae (30.6)–(30.9) are called Maxwell's relations.

The second and third of these relations are particularly important. They relate the derivatives of the entropy to the quantities $(\partial p/\partial T)_V$ and $(\partial V/\partial T)_p$ which can be measured directly.

The thermodynamic potentials E and F , and H and G , are not independent of each other. It is easy to establish the relations between them if use is made of their definitions and the definition of the entropy. Thus, from (28.9) and (29.7) we find

$$E = F + TS = F - T \left(\frac{\partial F}{\partial T} \right)_V. \quad (30.10)$$

Analogously,

$$H = G - T \left(\frac{\partial G}{\partial T} \right)_p. \quad (30.11)$$

Formulae (30.10) and (30.11) are called the Gibbs–Helmholtz equations.

The Gibbs–Helmholtz equations can also be written in the forms

$$\left(\frac{\partial(F/T)}{\partial T} \right)_V = - \frac{E}{T^2}, \quad (30.12)$$

$$\left(\frac{\partial(G/T)}{\partial T} \right)_p = - \frac{H}{T^2}. \quad (30.13)$$

If the dependences of the energy and enthalpy on the temperature are known, the integration of the Gibbs–Helmholtz equations allows one to find the dependence of the free energy and thermodynamic potential on the temperature:

$$F = -T \int \frac{E}{T^2} dT + \text{const} \cdot T, \quad (30.14)$$

$$G = -T \int \frac{H}{T^2} dT + \text{const} \cdot T. \quad (30.15)$$

§31. Methods for the transformation of thermodynamic quantities

In thermodynamics one often has to carry out transformations of thermodynamic quantities, for example, the transformations of variables or the replacement of certain quantities maintained constant in the course of a process by other quantities. Such transformations must be carried out according to the general rules for the substitution of variables when differentiating with respect to several variables. We shall present here one of the methods for such transformations*.

* For footnote see next page.

Let three variable quantities x, y, z be given such that each of them can be considered to be a single-valued function of the other two variables, i.e.

$$z = z(x, y), \quad y = y(x, z), \quad x = x(y, z).$$

We shall find the relation between the derivatives $(\partial z / \partial x)_y$ and $(\partial z / \partial y)_x$. For this we write the obvious equations

$$dz(x, y) = \left(\frac{\partial z}{\partial x} \right)_y dx + \left(\frac{\partial z}{\partial y} \right)_x dy, \quad (31.1)$$

$$dx = \left(\frac{\partial x}{\partial y} \right)_z dy + \left(\frac{\partial x}{\partial z} \right)_y dz.$$

Substituting dx from the lower equation into the upper one, we have

$$\begin{aligned} dz &= \left(\frac{\partial z}{\partial x} \right)_y \left(\frac{\partial x}{\partial y} \right)_z dy + \left(\frac{\partial z}{\partial x} \right)_y \left(\frac{\partial x}{\partial z} \right)_y dz + \left(\frac{\partial z}{\partial y} \right)_x dy = \\ &= \left[\left(\frac{\partial z}{\partial x} \right)_y \left(\frac{\partial x}{\partial y} \right)_z + \left(\frac{\partial z}{\partial y} \right)_x \right] dy + dz. \end{aligned} \quad (31.2)$$

Since dy is an arbitrary infinitesimal quantity, it is necessary for the fulfilment of (31.2) that

$$\left(\frac{\partial z}{\partial x} \right)_y \left(\frac{\partial x}{\partial y} \right)_z + \left(\frac{\partial z}{\partial y} \right)_x = 0.$$

Whence the relation sought follows:

$$\frac{(\partial z / \partial x)_y}{(\partial z / \partial y)_x} = - \left(\frac{\partial y}{\partial x} \right)_z. \quad (31.3)$$

Consider further the case when there are four quantities x, y, z, t , each pair of the quantities being completely determined if the other pair is defined, i.e.

* For another method, based on the use of the properties of Jacobians, see L.D. Landau and E.M. Lifshitz, *Course of theoretical physics*, Vol. 5, *Statistical Physics* (Pergamon, London, 1958).

$$t = t(x, y) = t(y, z) = t(x, z)$$

and so on.

Writing t as a function of the pair of variables x and y , we have

$$dt = \left(\frac{\partial t}{\partial x} \right)_y dx + \left(\frac{\partial t}{\partial y} \right)_x dy. \quad (31.4)$$

The same change in the quantity t as a function of y and z can be written in the form

$$dt = \left(\frac{\partial t}{\partial y} \right)_z dy + \left(\frac{\partial t}{\partial z} \right)_y dz. \quad (31.5)$$

Substituting the expression for dx from (31.1) into (31.4), we get

$$\begin{aligned} dt &= \left[\left(\frac{\partial t}{\partial x} \right)_y \left(\frac{\partial x}{\partial y} \right)_z + \left(\frac{\partial t}{\partial y} \right)_x \right] dy + \left(\frac{\partial t}{\partial x} \right)_y \left(\frac{\partial x}{\partial z} \right)_y dz = \\ &= \left[\left(\frac{\partial t}{\partial x} \right)_y \left(\frac{\partial x}{\partial y} \right)_z + \left(\frac{\partial t}{\partial y} \right)_x \right] dy + \left(\frac{\partial t}{\partial z} \right)_y dz. \end{aligned} \quad (31.6)$$

Comparing (31.5) and (31.6), we find

$$\left(\frac{\partial t}{\partial y} \right)_z = \left(\frac{\partial t}{\partial y} \right)_x + \left(\frac{\partial t}{\partial x} \right)_y \left(\frac{\partial x}{\partial y} \right)_z. \quad (31.7)$$

We shall present several examples of the use of the relations (31.3) and (31.7).

1. Find the relation between the derivatives $(\partial S / \partial V)_T$ and $(\partial S / \partial T)_V$, and between the derivatives $(\partial S / \partial p)_T$ and $(\partial S / \partial T)_p$.

According to formula (31.3) we have

$$\left. \begin{aligned} \frac{(\partial S / \partial V)_T}{(\partial S / \partial T)_V} &= - \left(\frac{\partial T}{\partial V} \right)_S; \\ \frac{(\partial S / \partial p)_T}{(\partial S / \partial T)_p} &= - \left(\frac{\partial T}{\partial p} \right)_S. \end{aligned} \right\} \quad (31.8)$$

2. The following quantities are called thermal coefficients:

$$\alpha = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_p \quad \text{the coefficient of thermal expansion ,}$$

$$\beta = \frac{1}{p} \left(\frac{\partial p}{\partial T} \right)_V \quad \text{the thermal coefficient of pressure ,}$$

$$\kappa_T = - \frac{1}{V} \left(\frac{\partial V}{\partial p} \right)_T \quad \text{the isothermal compressibility .}$$

Find the relation between them.

From the definitions of α and β and (31.3) it follows that

$$\frac{\alpha}{\kappa_T} = - \frac{(\partial V / \partial T)_p}{(\partial V / \partial p)_T} = \left(\frac{\partial p}{\partial T} \right)_V = p\beta . \quad (31.9)$$

3. Find the ratio of the adiabatic compressibility to the isothermal compressibility, expressing it in terms of heat capacities.

The adiabatic compressibility is defined as

$$\kappa_S = - \frac{1}{V} \left(\frac{\partial V}{\partial p} \right)_S . \quad (31.10)$$

Analogously, the isothermal compressibility is

$$\kappa_T = - \frac{1}{V} \left(\frac{\partial V}{\partial p} \right)_T . \quad (31.11)$$

According to formula (31.3) we find

$$\left(\frac{\partial V}{\partial p} \right)_S = - \frac{(\partial S / \partial p)_V}{(\partial S / \partial V)_p} ; \quad \left(\frac{\partial V}{\partial p} \right)_T = - \frac{(\partial T / \partial p)_V}{(\partial T / \partial V)_p} .$$

Dividing the first equality by the second one, we have

$$\frac{\kappa_S}{\kappa_T} = \frac{(\partial V / \partial p)_S}{(\partial V / \partial p)_T} = \frac{(\partial S / \partial p)_V}{(\partial T / \partial p)_V} \frac{(\partial T / \partial V)_p}{(\partial S / \partial V)_p} = \frac{(\partial S / \partial T)_V}{(\partial S / \partial T)_p} = \frac{C_V}{C_p} . \quad (31.12)$$

4. Find the relation between $(\partial H/\partial p)_T$ and $(\partial H/\partial T)_p$.

According to formula (31.3) we have

$$\frac{(\partial H/\partial p)_T}{(\partial H/\partial T)_p} = -\left(\frac{\partial T}{\partial p}\right)_H. \quad (31.13)$$

5. Find the relation between the heat capacities $C_p = T(\partial S/\partial T)_p$ and $C_V = T(\partial S/\partial T)_V$.

Since the relation between four quantities S , T , p and V is to be established, use should be made of formula (31.7). It gives

$$\left(\frac{\partial S}{\partial T}\right)_p = \left(\frac{\partial S}{\partial T}\right)_V + \left(\frac{\partial S}{\partial V}\right)_T \left(\frac{\partial V}{\partial T}\right)_p.$$

From the Maxwell relation (30.7) and the definition of the heat capacities we find

$$C_p = C_V + T\left(\frac{\partial p}{\partial T}\right)_V \left(\frac{\partial V}{\partial T}\right)_p. \quad (31.14)$$

By means of (31.9) we can write

$$C_p = C_V - \frac{T(\partial V/\partial T)_p^2}{(\partial V/\partial p)_T} = C_V + \frac{\alpha^2 TV}{\kappa_T}. \quad (31.15)$$

6. Find the relation between the isothermal compressibility $\kappa_T = -V^{-1}(\partial V/\partial p)_T$ and the adiabatic compressibility $\kappa_S = -V^{-1}(\partial V/\partial p)_S$.

Making use of (31.7), we have

$$\left(\frac{\partial V}{\partial p}\right)_T = \left(\frac{\partial V}{\partial p}\right)_S + \left(\frac{\partial V}{\partial S}\right)_p \left(\frac{\partial S}{\partial p}\right)_T.$$

By means of formula (30.8) and the equality

$$\left(\frac{\partial S}{\partial V}\right)_p = \left(\frac{\partial S}{\partial T}\right)_p \left(\frac{\partial T}{\partial V}\right)_p$$

we obtain

$$\left(\frac{\partial V}{\partial p}\right)_T = \left(\frac{\partial V}{\partial p}\right)_S - \frac{T}{C_p} \left(\frac{\partial V}{\partial T}\right)_p^2 = \left(\frac{\partial V}{\partial p}\right)_S - \frac{TV^2\alpha^2}{C_p},$$

or

$$\kappa_T = \kappa_S + \frac{TV\alpha^2}{C_p}. \quad (31.16)$$

§ 32. The determination of thermodynamic quantities by the methods of statistical physics

For the determination of the thermodynamic quantities which we have introduced there are two possibilities:

- (1) to calculate them by the methods of statistical physics,
- (2) to find them on the basis of certain thermal measurements.

We shall begin with the analysis of methods of calculating thermodynamic quantities.

As can be seen from formula (21.3), the internal energy of any body can be found if its partition function Z is known. According to (24.8), it is also necessary to calculate the partition function to find the entropy.

Substituting the value of S using (24.8) into the definition of the free energy, we have

$$F = E - TS = -kT \ln Z. \quad (32.1)$$

The expression of F in terms of Z turns out to be particularly simple.

In addition we write an explicit expression for the pressure. According to (22.5) and (22.3) we can write

$$-p\delta V = \sum (w_i \delta \epsilon_i)_{w_i},$$

hence

$$p = -\frac{1}{Z} \sum \frac{\partial \epsilon_i}{\partial V} e^{-\epsilon_i/\theta} \Omega(\epsilon_i). \quad (32.2)$$

We have made use of the fact that the probability distribution remains unchanged, and have put in its explicit expression. Formula (32.2) can be re-

written in a standard form, expressing it in terms of the partition function Z . For this we note that

$$\frac{\partial Z}{\partial V} = \frac{\partial}{\partial V} \sum e^{-\epsilon_i/\theta} \Omega(\epsilon_i) = -\frac{1}{\theta} \sum e^{-\epsilon_i/\theta} \frac{\partial \epsilon_i}{\partial V} \Omega(\epsilon_i) \quad (32.3)$$

and, consequently,

$$\frac{\partial \ln Z}{\partial V} = \frac{1}{Z} \frac{\partial Z}{\partial V} = -\frac{1}{\theta Z} \sum \frac{\partial \epsilon_i}{\partial V} e^{-\epsilon_i/\theta} \Omega(\epsilon_i). \quad (32.4)$$

Comparing (32.4) and (32.2) we see that the pressure can be written in the form

$$p = \theta \frac{\partial \ln Z}{\partial V} = kT \frac{\partial \ln Z}{\partial V}. \quad (32.5)$$

Comparing the expression (32.5) with the expression (32.1) for the free energy, we find, in correspondence with (29.8),

$$p = -\left(\frac{\partial F}{\partial V}\right)_T. \quad (32.6)$$

From formula (32.6) it follows that, knowing the partition function Z of the system, one can find the equation of state. Indeed, since $F = f(V, T)$, formula (32.6) establishes the relation between the pressure, volume and temperature of the system.

From the method of proof it can be seen that the mean value of the derivative $\partial \epsilon_i / \partial V$, i.e. $\overline{\partial \epsilon_i / \partial V}$, cannot be identified with the derivative of the mean energy with respect to the volume $\partial \bar{\epsilon} / \partial V$, i.e. with $\partial E / \partial V$. The latter is by virtue of (30.10) and (29.8) obviously equal to

$$\frac{\partial E}{\partial V} = \frac{\partial}{\partial V} \left(F - T \frac{\partial F}{\partial T} \right) = -p + T \frac{\partial p}{\partial T}.$$

The Gibbs thermodynamic potential G will play an important role in further developments. We shall find its statistical expression. From the definition of the thermodynamic potential it follows that, in contrast to the free energy, it is a function of the pressure and not of the volume. In other words, G is a function of the generalized force $\Lambda = p$ and not of the generalized

coordinate $\lambda = V$. It is with this fact that the important role of G in thermodynamics is associated: in practice it is simpler to maintain a constant value of the pressure and other generalized forces than to maintain a constant value of the volume and corresponding generalized coordinates.

To obtain a statistical expression for G it is necessary to find the dependence of the partition function Z on the generalized force (the pressure p) as an independent variable, i.e.

$$Z = Z(T, p) .$$

In the course of the foregoing discussion we have assumed the energy levels of the system to depend on the parameter λ , i.e. $\epsilon_i = \epsilon_i(\lambda)$. We shall now assume the force Λ to be a variable, and λ to be a function of Λ .

Consider as an example an ideal gas in a container having a movable wall. If the independent variable is an external parameter (the volume of the container), then the system is the gas in the container. But if the independent variable is an acting force (an external pressure) then the movable wall should be included as a part of the system.

Thus, the subsystem is formed by N gas molecules and the movable wall of the container, so that the subsystem has altogether $3N + 1$ degrees of freedom. Its state is characterized by the coordinates and momenta of all the molecules, as well as by the position and momentum of the movable wall.

The movable wall is acted upon by a pressure p . A change in the pressure will lead to a change in the volume of the system. The latter in its turn leads to a shift of the energy levels of the system. We write the energy of the system (a gas + a movable wall) in the form

$$\epsilon_i(p) = \epsilon_i + (\epsilon_{\text{kin}} + \epsilon_{\text{pot}}) ,$$

where ϵ_i is the energy of the gas, and $(\epsilon_{\text{kin}} + \epsilon_{\text{pot}})$ is the energy of the wall.

To find the potential energy ϵ_{pot} of the wall we note that the work produced on the system as the external parameter p changes by an amount δp is equal to

$$\delta W_p = -V \delta p .$$

Hence for ϵ_{pot} we can write

$$\epsilon_{\text{pot}} = Vp .$$

The kinetic energy of the thermal motion of the wall ϵ_{kin} can be disregarded in comparison with the kinetic energy of the gas molecules (since the number of the latter is very large). Hence finally we have

$$\epsilon_i(p) = \epsilon_i + pV.$$

The partition function of the system has the form

$$Z(p, T) = \int \sum_i \exp\left(-\frac{\epsilon_i + pV}{kT}\right) \Omega(\epsilon_i) dV, \quad (32.7)$$

where the summation is carried out over all levels of the system (the value of ϵ_i depends on V), while the integration is carried out over the entire volume of the system. In analogy with (32.1) we can write

$$G = -kT \ln Z(p, T). \quad (32.8)$$

Formula (32.8) shows that the logarithm of the partition function represents the free energy in the broad sense: F in the case of the variable V , and G in the case of the variable p . Since neither the concrete nature of the system nor the character of the generalized force have figured in the foregoing reasoning, the expression obtained for G remains valid for any system and for any generalized force.

From formula (32.8) one can find the mean volume of the system:

$$V = \left(\frac{\partial G}{\partial p}\right)_T = -kT \frac{\partial \ln Z(p, T)}{\partial p}. \quad (32.9)$$

The relations obtained allow one to express thermodynamic functions — the internal energy of a body, its free energy, entropy and pressure — directly in terms of the partition function Z . The value of the partition function is determined by the molecular properties of the system — its possible energy states — as well as by the temperature T and volume or pressure. Thus, we arrive at the important conclusion:

Statistical physics allows one to find the values of thermodynamic quantities in a purely mathematical way, provided the energy levels of the system and the statistical weights of the corresponding states are known.

However, the role and significance of statistical physics do not reduce to this very important but nevertheless partial result. Statistical physics allows one to attach a more profound meaning to thermodynamic quantities and

ideas, and reveals the physical laws underlying the thermodynamic behaviour of a system. Thus, we have seen that the ideas of thermodynamic energy, work, entropy and amount of heat were given a clear physical interpretation. All these concepts have been associated with molecular processes taking place in the system. The general statement that heat is a form of motion has found its mathematical expression in the formulae obtained.

True, up to now we have not considered real physical systems and have not detailed the character of the molecular motion and quantum states of the system. The following chapters, in which the general laws of statistical physics will be applied to different physical systems, are devoted to the application of the laws found.

The general character of our formulation of statistical laws has very important advantages. Namely, owing to their general character, statistical laws, which were originally found for systems obeying the laws of classical mechanics, are still valid, and have undergone only small modifications associated with the replacement of classical systems by quantum systems. The general character of the laws of statistical mechanics do not restrict the sphere of its considerations to purely thermal processes, which constitute the primary basis of statistics, but include also the most diverse properties of matter — electric, magnetic, chemical and so on. These properties of matter will be considered in Part IV of the book. In the meanwhile we shall confine ourselves to the study of the thermal properties of matter.

§ 33. The determination of thermodynamic quantities from experimental data

We shall now consider the general methods of finding thermodynamic quantities from experimental data. Before passing on to thermodynamic potentials, we shall discuss the problem of establishing a scale of absolute temperature. To obtain this scale it is necessary to find a relationship between the absolute temperature and the empirical temperature measured on an arbitrary scale by any thermometer.

We find the relation between the quantities $(\partial E/\partial V)_T$ and $(\partial E/\partial V)_S$, making use of formula (31.7):

$$\left(\frac{\partial E}{\partial V}\right)_T = \left(\frac{\partial E}{\partial V}\right)_S + \left(\frac{\partial E}{\partial S}\right)_V \left(\frac{\partial S}{\partial V}\right)_T.$$

Making use of (24.6) and (30.7), we have

$$\left(\frac{\partial E}{\partial V}\right)_T = -p + T\left(\frac{\partial p}{\partial T}\right)_V.$$

Now let the temperature be measured on an arbitrary empirical scale of temperatures t , for which one can write that

$$t = f(T) \quad \text{or} \quad T = \varphi(t),$$

where f is an unknown function. Then, obviously

$$\left(\frac{\partial E}{\partial V}\right)_T = \left(\frac{\partial E}{\partial V}\right)_t, \quad \left(\frac{\partial p}{\partial T}\right)_V = \left(\frac{\partial p}{\partial t}\right)_V \frac{dt}{dT}.$$

Consequently,

$$\left(\frac{\partial E}{\partial V}\right)_t = -p + \left(\frac{\partial p}{\partial t}\right)_V \frac{dt}{d \ln T},$$

hence

$$\frac{d \ln T}{dt} = \frac{(\partial p / \partial t)_V}{(\partial E / \partial V)_t + p} = \psi(t, V). \quad (33.1)$$

Only quantities which can be measured directly and expressed in terms of the volume V and empirical temperature t stand on the right-hand side of eq. (33.1). Hence the dependence of T on t can be found by integrating (33.1) upon determining the function $\psi(t, V)$ experimentally. In fact, for the establishment of the absolute temperature scale, for example, at temperatures close to the absolute zero or at very high temperatures, use is made not only of thermal methods of measurement but also of magnetic or optical ones.

We now pass on to methods of finding thermodynamic quantities. The determination of the enthalpy is the simplest.

At constant pressure

$$dH = TdS = C_p dT.$$

On integrating we have

$$H = \int_{T_1}^T C_p(T) dT + H_1. \quad (33.2)$$

In order to find $H(T)$ experimentally it is necessary to measure the heat capacity C_p in the entire temperature range and to know H_1 at a certain temperature. If, as is usually the case, we are interested in the change of the heat content, then

$$\Delta H = H(T) - H(T_1) = \int_{T_1}^T C_p(T) dT$$

and the constant H_1 drops out.

For the absolute measurement of H it is necessary to make use of the properties of H at absolute zero, which are established by the third law of thermodynamics (see §35 below).

In melting and evaporation the enthalpy increases at a fixed temperature (see §62).

The entropy is found from the values of the heat capacity:

$$S = \int_{T_1}^T \frac{C_p dT}{T} + S_1 = \int_{T_1}^T C_p d \ln T + S_1.$$

The value of S is obtained by the graphical integration of the curve $C_p = f(\ln T)$. The change in the entropy is

$$\Delta S = S(T) - S(T_1) = \int_{T_1}^T C_p d \ln T. \quad (33.3)$$

The value of $S(0)$ is determined by the third law of thermodynamics (see §35).

To find the entropy of a liquid and vapour it is necessary to take into account the change in the entropy as the latent heat of melting and vaporization is absorbed (see §62).

The dependence of the entropy on the pressure is determined from formula (30.8):

$$\left(\frac{\partial S}{\partial p} \right)_T = - \left(\frac{\partial V}{\partial T} \right)_p = -\alpha V.$$

Integrating, we find

$$S(T, p) = - \int_{p_1}^p \alpha V dp + S(T, p_1) \quad (33.4)$$

or, for the change in the entropy,

$$S(T, p) - S(T, p_1) = - \int_{p_1}^p \alpha V dp. \quad (33.5)$$

Knowing the equation of state $V = V(T, p)$ one can find the coefficient of thermal expansion α and, by means of formula (33.5), the dependence of the entropy on the pressure.

Knowing H , S and the equation of state one can find all the other thermodynamic potentials:

$$E = H - pV, \quad F = E - TS, \quad G = F + pV.$$

Thus, in order to find all the thermodynamic potentials of a given substance it is necessary:

- (1) to measure the heat capacity over the entire temperature interval of interest,
- (2) to measure the coefficient of thermal expansion, and
- (3) to determine the equation of state.

H , E and S are determined to within constants representing the value of these quantities at a certain temperature, while F and G are determined to within a linear function of the temperature.

In conclusion we should discuss the thermodynamic potentials of chemical compounds. As a chemical compound is formed changes in the enthalpy ΔH and the entropy $\Delta S = T^{-1} \Delta H$ take place (§67). Correspondingly, the thermodynamic potentials of chemical compounds are the sum of their values for the initial substances and their changes in the course of the reaction.

The introduction of thermodynamic potentials allows one, in principle, to determine the thermodynamic potentials of complex substances, if the thermodynamic potentials of the elements constituting them and the heats of the chemical reactions of formation of the compounds are known.

§34. Throttling

The process in which a gas initially occupying a volume V_1 at a constant pressure p_1 passes out of its container into a container with a volume V_2 at a constant pressure p_2 , plays an important role in contemporary technology. This process is called the throttling process or the Joule–Thompson process.

The throttling process is one of the basic methods of obtaining low temperatures in contemporary cryogenic techniques. In practice the throttling process is carried out by slowly forcing the gas from one container into another through a system of thin capillary tubes offering a great hydrodynamic impedance to the flow of the gas. This great hydrodynamic impedance ensures a small macroscopic velocity of motion of the gas. The forcing through of the gas is carried out in adiabatic conditions, for which the apparatus is enclosed by thermally insulating material.

Since no heat is supplied to the system, and the energy dissipation due to friction can be disregarded (in view of the small velocity of motion of the gas), the change in the internal energy of the gas ΔE is equal to the mechanical work done on the gas:

$$\Delta E = E_2 - E_1 = W,$$

where E_1 and E_2 are respectively the energy of the gas in the initial and final states. The latter is made up of the work of compression done on the gas at a pressure p_1 (from the initial volume V_1 to the final volume equal to zero) and the work of expansion produced by the gas at a pressure p_2 (from the initial volume equal to zero to the final volume V_2), i.e.

$$W = -\left(\int_{V_1}^0 p_1 dV + \int_0^{V_2} p_2 dV \right) = -(p_2 V_2 - p_1 V_1).$$

Hence for the throttling process one can write that

$$E_1 + p_1 V_1 = E_2 + p_2 V_2,$$

or

$$H_1 = H_2. \quad (34.1)$$

Thus, the throttling process represents a process at constant enthalpy of the gas.

Assuming p_1 and p_2 to be very close to each other (and $p_2 < p_1$, $\Delta p < 0$), we find the change in the temperature of the gas as the pressure changes in the throttling process. This change is characterized by the derivative $\mu = (\partial T / \partial p)_H$ called the Joule–Thompson coefficient. The value of the latter has been found above (see (31.13)). We write (31.13) in the form

$$\begin{aligned}
 \mu &= \left(\frac{\partial T}{\partial p} \right)_H = - \frac{(\partial H / \partial p)_T}{(\partial H / \partial T)_p} = - \frac{(\partial (G + TS) / \partial p)_T}{C_p} \\
 &= - \frac{V + T(\partial S / \partial p)_T}{C_p} = - \frac{-T(\partial V / \partial T)_p + V}{C_p}. \quad (34.2)
 \end{aligned}$$

We apply formula (34.2) first of all to an ideal gas. In this case, obviously,

$$\mu = \left(\frac{\partial T}{\partial p} \right)_H = 0.$$

When an ideal gas is throttled no change in its temperature takes place. The meaning of this is obvious: the internal energy of an ideal gas does not depend on the volume and does not change as the gas expands. The situation is different in the case of a real gas for which the energy of interaction between molecules and, consequently, the internal energy, depends on the volume. The theory of a real gas will be expounded in §46–47.

In the approximation of eq. (46.18) we find

$$\mu = - \frac{-T(dV/dT)_p + V}{C_p} \approx \frac{N[\beta - T(d\beta/dT)]}{2C_p},$$

where β is the constant contained in the Van der Waals equation. Since $C_p > 0$, the sign of μ is determined by that of the numerator. From the definition (46.14) of β we find

$$\beta - T \frac{d\beta}{dT} = \int \left[e^{-u(r)/kT} \left(1 - \frac{u(r)}{kT} \right) - 1 \right] dV. \quad (34.3)$$

At high temperatures in the region of attraction $r > d$, $u(r) \ll kT$ and $e^{-u/kT} \approx 1$. In the region of repulsion $r < d$, $u(r) \gg kT$ and $e^{-u/kT} \ll 1$. In this case the entire integrand in the expression (34.3) is negative, and $\beta - T(d\beta/dT) < 0$. At low temperatures in the region of attraction $r > d$, $|u(r)| \gg kT$ and $u(r)/kT \gg 1$. In this case the integrand is positive, and $\beta - T(d\beta/dT) > 0$.

Thus, in throttling at high temperatures $\mu < 0$ and the temperature of the gas increases, whereas in throttling at low temperatures $\mu > 0$ and the temperature of the gas decreases. At a certain temperature, called the inversion point of a given gas, its coefficient $\mu = 0$. Carrying out the throttling at temperatures lower than the inversion point, one can cool gases to very low temperatures. Throttling is one of the most widely used methods of cooling gases.

§35. The third law of thermodynamics

Consider the behaviour of a certain macroscopic (thermodynamic) system at very low temperatures. We assume that the system is in a state of statistical equilibrium with an energy $\bar{\epsilon}$, so that its entropy is determined by the Boltzmann formula. Let the possible values of the energy of the system (its energy levels) form a sequence $\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3, \dots$, where ϵ_0 is the lowest possible energy (the normal level of the system), and $\epsilon_1, \epsilon_2, \epsilon_3, \dots$, are excited energy levels. The energy levels get closer together very rapidly as the excitation energy increases. However, the fact that the spacing $\epsilon_1 - \epsilon_0 = \Delta\epsilon$ between the normal level and the first excited level is a finite, if only an extremely small quantity, is of great importance.

If the temperature of the system is sufficiently low, so that the thermal energy, kT , is considerably smaller than the spacing between the lowest level and the first excited level, i.e. $kT \ll \Delta\epsilon$, then the thermal excitations of the system are insufficient for the system to get into the state ϵ_1 . Hence at a very low temperature the system must be in the state with the lowest energy ϵ_0 . The thermodynamic energy of the system is obviously equal to

$$E_0 = \epsilon_0, \quad (35.1)$$

and does not depend on the temperature (for $T \ll k^{-1}\Delta\epsilon$). Hence it follows that the heat capacity of the system at constant volume is

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = \left(\frac{\partial \epsilon_0}{\partial T} \right)_V = 0. \quad (35.2)$$

Now we find the entropy of the system. According to the Boltzmann formula the entropy is equal to

$$S = k \ln \Omega_0, \quad (35.3)$$

where Ω_0 is the number of states of the system with the energy ϵ_0 . But at absolute zero an equilibrium system is in a quite definite state the energy of which is exactly equal to E_0 . We know, however, that if the energy of the system is exactly defined, then by this very fact the state of the system is also defined. Hence the number of states with energy ϵ_0 is simply equal to unity*. Then from formula (35.3) it follows that the entropy of the system at absolute zero is equal to zero:

* For footnote see next page.

$$S = 0 \quad \text{as} \quad T \rightarrow 0. \quad (35.4)$$

The condition (35.4) was first established by Nernst, and is called the third law of thermodynamics, or the Nernst heat theorem. It was introduced, not statistically, but on the basis of an analysis of experimental data, in particular, the thermal effects of chemical reactions at low temperatures.

The condition (35.4) is not a consequence of the first and second laws of thermodynamics. Its role in contemporary thermodynamics is very important.

As we have pointed out already in §33, in order to determine the values of thermodynamic potentials from empirical or statistical data it is necessary to know their value at absolute zero. The third law of thermodynamics allows one to do this.

It should be stressed that the third law of thermodynamics is closely associated with the quantum character of the system. If the system considered obeyed the laws of classical mechanics, then its energy would change continuously. Hence, however low the temperature T may be, the energy of thermal excitation kT of the system would be infinitely large in comparison with the infinitesimal small spacing between energy levels in the classical system. To a finite energy interval kT there would correspond an infinitely large number of possible states Ω . The entropy would be large at as low (but finite) a temperature as one wished.

The quantum character of real systems, which is shown very weakly at high temperatures, assumes, as we see, paramount importance at very low temperatures. This is in complete agreement with the general propositions stated in §1.

In addition a few words should be said apropos the constant entropy. Although we write $S = 0$ as $T \rightarrow 0$, in fact formula (35.4) should be written in the form (see §24)

$$S \rightarrow \text{const} \quad \text{as} \quad T \rightarrow 0. \quad (35.5)$$

The constant in formula (35.5) cannot be determined, since it is an arbitrary integration constant.

The value of the constant in (35.5) does not depend on the pressure, volume and other parameters characterizing the state of the system. Whatever the state in which the substance may be – in the form of a chemical com-

* If the state with energy ϵ_0 is accidentally degenerate and the energy ϵ_0 is possessed by several states, then the situation is not changed. If $\Omega_0(\epsilon_0)$ is not a large number, then $k \ln \Omega_0(\epsilon_0)$ is practically equal to zero, in view of the extreme smallness of k .

pound or a pure element, with a high or low density and so on — the value of this constant will be the same.

The difference between the entropies of two thermodynamic states of a system, which differ by different values of the parameters, tends to zero as $T \rightarrow 0$. Thus, for example, the entropy S_1 of a mixture of two moles of elements A and B is equal to the entropy S_2 of one mole of their compound AB as $T \rightarrow 0$.

Such a statement follows from those experimental data which established the third law of thermodynamics. One rather frequently encounters the statement that by means of the third law of thermodynamics the integration constant in the formula for the entropy can be determined and that by this very fact its absolute value can be determined. The reasoning of this paragraph and §24 shows clearly that the absolute value of the entropy has no physical meaning: the entropy is determined in essence only to within an arbitrary constant.

The meaning of the third law does not lie in the fact that it allows one to find the absolute value of the entropy but in the fact that it establishes the constancy of the entropy as $T \rightarrow 0$ (in the sense of the independence from the parameters which characterize the state). The value of this constant can be chosen as zero entropy.

The third law of thermodynamics is often formulated as the principle of the unattainability of absolute zero. Such a formulation follows from condition (35.4). If the entropy of the system is $S = 0$, then by means of it and a cycle of two adiabatic lines $S = S_1$ and $S = S_2$ connected by the isothermal line $T = 0$ (on which $S = 0$) and another arbitrary isothermal line a perpetual motion machine of the second kind can be constructed. However, it should be noted that the unattainability of absolute zero does not contradict the possibility of obtaining in principle temperatures differing from $T = 0$ by as little as one wishes.

In the foregoing reasoning it has been assumed that the system is in an equilibrium state. However, we have not made any assumptions concerning its state of aggregation. The expression (35.4) must be equally valid for solid, liquid or gaseous systems.

There exists only one system which remains liquid near absolute zero: helium II. As to gases, all ordinary gases at a sufficient pressure are condensed long before those low temperatures at which the entropy tends to zero. Hence the pressure of the saturated vapour above the solid body is negligible as $T \rightarrow 0$. However, there exist systems which can conditionally be assumed to be gaseous as $T \rightarrow 0$. There is, first of all, the electron gas in metals, the properties of which will be considered in more detail in ch. 7 of Part IV.

Table 1
The entropy of HCl from experimental calorimetric data

State of aggregation	Temperature interval (K)	Method of determination	Entropy ($\text{J mol}^{-1} \text{K}^{-1}$)
Solid	0–16	Extrapolation of the curve obtained at higher temperatures	1.26
	16–98.36	From measured values of the heat capacity	29.6
	98.36	Latent heat	12.1
	Phase transition into another crystal modification		
	98.36–158.91	From the heat capacity	21.1
	158.91	Latent heat	12.6
Liquid	Melting		
	158.91–188.07	From the heat capacity	9.84
	188.07	Latent heat	85.89
	Boiling		total 172.3
Gaseous	188.17	Measured entropy	172.5
	188.17	Entropy calculated theoretically with the correction for the fact that the gas is not an ideal gas	172.9

However, all ordinary substances pass over into the solid state at a sufficiently low temperature. It should be noted that, in addition to true solid bodies, i.e. crystals, a rather large number of substances are in the solid, or amorphous, state. Although amorphous bodies can possess a number of properties very similar to those of true solid bodies, they in reality represent supercooled liquids. We shall dwell on the properties of supercooled liquids somewhat later.

The validity of the third law of thermodynamics was checked experimentally by different methods for a large number of substances. Although the experimental data confirming the validity of the third law of thermodynamics are not yet so numerous and versatile as those confirming the second law, they remove all doubts as to its validity. The most accurate method of checking it is the investigation of chemical equilibria at low temperatures.

For a number of substances direct measurements of the heat capacity have been carried out, on the basis of which the values of the entropy have been found. The latter were compared with those calculated theoretically on the basis of the assumption that $S = 0$ as $T \rightarrow 0$. Table 1 is an example of such a direct check.

Finally, an experimental verification of certain consequences of the third law of thermodynamics has been carried out for many substances. Thus, for example, from the condition (35.4) it follows also that

$$\left(\frac{\partial S}{\partial p}\right)_T = 0 \quad \text{as} \quad T \rightarrow 0.$$

But

$$\left(\frac{\partial S}{\partial p}\right)_T = -\frac{\partial^2 G}{\partial p \partial T} = -\left(\frac{\partial V}{\partial T}\right)_p.$$

Hence from the third law of thermodynamics it follows that the coefficient of thermal expansion must reduce to zero as $T \rightarrow 0$. The measurements of this coefficient for a number of crystals (diamond, HCl, Cu and so on) confirm this conclusion.

However, for a number of substances the measurements of the entropy showed the condition $S = 0$ as $T = 0$ to be unfulfilled. The number of such substances is relatively large. They are amorphous bodies, alloys, and a number of chemical compounds: CO, NO, H₂O and others. The existence of so many exceptions to the third law led to doubts as to its general applicability. However the statistical interpretation of the third law allowed one to explain the true origin of these violations.

In deriving the third law we have assumed that our system at absolute zero is in an equilibrium state and that it goes over into the state with the lowest energy ϵ_0 as the temperature tends to zero. However, there are systems which are not in an equilibrium state as $T \approx 0$, to which the third law of thermodynamics is inapplicable. These systems possess an entropy different from zero at absolute zero.

We have already said that the relaxation time can vary within an extremely large range, and that in a number of cases it can have very large values. This holds particularly at low temperatures, when the thermal energy is not large. In this case the establishment of an equilibrium state proceeds particularly slowly. For concreteness we imagine that our system is made of diatomic molecules (for example, CO molecules) forming a regular crystal lattice. In



Fig. III.13

fig. III.13 each molecule is presented schematically by an arrow one end of which represents a carbon atom and whose other end represents an oxygen atom. Two types of orientation of CO molecules in the lattice are possible, shown schematically in figs. III.13a and b. In one case there is a completely random orientation of the molecules, and in the other case there is a regular orientation of the molecules. All properties of carbon atoms are so similar to those of oxygen atoms that there is a negligible difference between the two forms of the crystal, with regularly and chaotically oriented molecules. Their symmetry and basic properties are completely identical. The difference between the energies of the two states is very small. The state with the regular orientation of the molecules corresponds to total equilibrium as $T \rightarrow 0$, for the energy of this turns out to be somewhat lower than that of the state with the random orientation. However, at higher temperatures the random orientation of the molecules becomes the equilibrium orientation. CO crystallizes at a relatively high temperature, when the random orientation is the equilibrium orientation. As the temperature is lowered down to $T \ll k^{-1} \Delta \epsilon$ the random orientation becomes non-equilibrium, so that the molecules should pass over into the state with the regular orientation. However, the transition of the molecules from a random distribution to a regular distribution at low temperatures proceeds so slowly that even for reasonably large observation times total equilibrium does not have sufficient time to be established. The system, at as low a temperature as one wishes, will be in a state with a random orientation of the molecules. This state will be the one with the lowest realizable energy. It is not an equilibrium state, since for a regular orientation the energy of the system would be even lower, but at the same time at low temperatures, when the velocity of the process of orientation becomes very small, the system can be in this state for an extremely long time (practically as long as one wants). The state with a random distribution of molecular orientations is, thus, a metastable state. The reasoning presented above is inapplicable to a system in a metastable state, and hence its entropy does not satisfy the condition (35.4).

Finding the entropy of a system with a completely random arrangement of molecules presents no difficulty. Each molecule can, with equal proba-

bility be in two states differing only by their orientation. If the system contains N molecules, then the total number of states is obviously equal to $\Omega_0 = 2^N$ (in view of the simplicity of the problem there is no need to express this number in terms of the phase volume). The entropy of the system at absolute zero is equal to

$$S_{T \rightarrow 0} = k \ln \Omega_0 = k \ln 2^N = Nk \ln 2 = 5.76 \text{ J mol}^{-1} \text{ K}^{-1}.$$

The measured value of the entropy of CO as $T \rightarrow 0$, turns out to be equal to $4.6 \text{ J mol}^{-1} \text{ K}^{-1}$. We see that the entropy of the system is indeed different from zero, its value being close to the theoretical one.

The third law of thermodynamics is of very great importance for finding the values of thermodynamic functions. In practical use it should, however, always be borne in mind that it refers only to systems in an equilibrium state, and that it is inapplicable to metastable systems. The example presented is not the only one. In § 54 we shall return to the problem of the behaviour of systems at low temperatures and shall present other examples of apparent deviations from the third law of thermodynamics. In any case, the possibility of applying the third law of thermodynamics is often not obvious and calls for great care. Before the appearance of the statistical derivation of the third law, when the limits of its applicability were still not understood, an inconsiderate application of the third law to metastable systems led to the contradictions mentioned above.

It should be noted in addition that, although the third law of thermodynamics is a very important proposition, the degree of its importance for science can scarcely be compared with that of the second law. In this sense the term "the third law" appears to be not quite apt.

§ 36. The statistical character of the second law of thermodynamics

In the preceding paragraphs we have established that the law of increase of entropy which in thermodynamics represents a direct generalization of experimental results, assumes a new, more profound, and at the same time clearer meaning in the light of the arguments of statistical physics. From the point of view of statistical physics the law of increase of entropy represents an expression of statistical laws shown in systems consisting of a very large number of molecules. In these systems, owing to the forces of intermolecular interaction, transitions from less probable states into more probable ones always take place until the system comes into the most probable state, i.e.

the state of total statistical equilibrium. This transition from a non-equilibrium state into an equilibrium state proceeds via complex processes in which an enormous number of molecules take part. The mechanism of the establishment of equilibrium and the character of the processes involved in it depend in many respects on the actual properties of the system.

The establishment of equilibrium (i.e. molecular disorder) in an ideal gas serves as the simplest example. Molecular disorder is established as a result of the collisions of molecules with the wall of the container and with one another.

It is not the way in which equilibrium is established in a system that will interest us now, but only the fact that equilibrium will, without fail, be established eventually. We have seen that the statistical formulation of the second law of thermodynamics differs from the thermodynamic one in a very important respect: in the former the following words are employed: "the most probable trend of processes", whereas in thermodynamics one speaks simply of the trend of processes. The formulation of statistical physics has a considerably less categorical character. It does not exclude at all, but, on the contrary, envisages the possibility of processes in the course of which the system passes over from a more probable state into a less probable one, and in which its entropy decreases. The existence of such processes, called fluctuations, is completely denied in the thermodynamic formulation of the second law.

For instance, imagine that a gas occupies one half of a free volume. According to the laws of thermodynamics the gas must expand and occupy the entire volume, the expansion being accompanied by an increase in the entropy. From the point of view of statistical physics such a behaviour of the gas is the most probable one. However, the possibility that the gas will not expand but compress is not excluded. In a macroscopic system the latter process has a probability which is negligibly small in comparison with that of the process of expansion. Hence in practice in macroscopic systems the first process will always occur.

Fluctuations occurring in a system which is already in an equilibrium state are of great importance. Imagine the gas occupying the entire volume with a uniform density and in an equilibrium state. If this gas is not subjected to an external action, then from the point of view of thermodynamics it will remain in this state for an indefinitely long time. Statistical physics states that, although for an overwhelmingly large part of the time the gas will occupy the entire volume and be in the equilibrium state, the possibility of fluctuations in the course of which the gas will spontaneously go out of the equilibrium state is not excluded. In particular, the gas can pass over spon-

taneously into a state in which it occupies only a part of the entire volume. The probability of such a transition is determined by the Boltzmann formula.

We shall not dwell now on the analysis of concrete examples of fluctuations, since ch. 8 is devoted to this important phenomenon. We stress only that experiment has completely confirmed the predictions of statistics as to the existence in nature of such spontaneous processes accompanied by a decrease in entropy. But the question naturally arises as to whether or not the statistical formulation of the second law contradicts the purely thermodynamic one. Does it follow from the statistical formulation of the second law that the construction of a perpetual motion machine of the second kind is a difficult but in principle soluble problem? For its solution can use be made of fluctuation processes accompanied by a decrease in the entropy? This problem was a subject of discussion over the course of a number of years, and its solution turned out to be very fruitful for the development of the basic propositions of statistical physics. But before giving an answer to this question it is necessary to discuss another, no less complicated question, logically preceding the former: in fact, how could it happen that, in considering molecular processes, we arrived at the idea of irreversibility?

It is well known that the laws of mechanics are strictly reversible. This is seen if only from the fact that the equation of classical mechanics

$$m \frac{d^2 r}{dt^2} = F$$

remains unchanged when the sign of the time is changed. A total reversibility reigns also in the realm of intra-atomic processes: from the laws of quantum mechanics there follows the principle of microscopic reversibility, stating that for all microscopic atomic or molecular processes the probability of direct processes is equal to that of reverse processes. Thus, there is no irreversibility in the basic laws of molecular processes. All the processes are strictly symmetric with respect to the future and past. On the other hand, statistical physics, based on molecular laws, leads to the appearance of irreversibility. At first sight it may seem that the laws of statistical physics contradict those of molecular motion, on the basis of which they were derived. In reality, however, this is not so. An increase in the entropy takes place when the system was initially in a certain non-equilibrium state. Then the most probable behaviour of the system in the course of time will be its transition into an equilibrium state. In fig. III.14 the time is plotted on the x -axis, and the entropy is plotted on the y -axis; the solid curve represents this most probable transition, while the dotted line represents the improbable

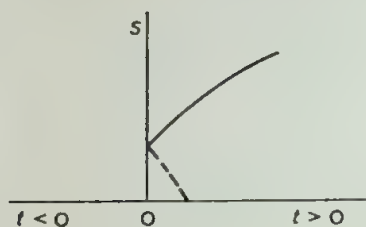


Fig. III.14

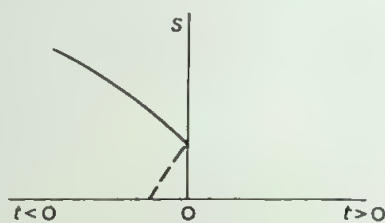


Fig. III.15

transition into a state with lower entropy *. Here the direction of time, the difference between the initial and final instants of time, appears completely obvious. However, if one thinks more carefully over this reasoning, then it can be seen that we always begin the consideration deliberately with a non-equilibrium state of the system. A certain asymmetry is concealed in the very statement of the problem. We say: in the beginning a non-equilibrium state is given; what will happen afterwards with the system left to itself? Let us try, however, to imagine how this initial non-equilibrium state could arise. It could arise either by an intervention in the system from without, or spontaneously in a closed system. In the first case it is clear that the absence of symmetry in the behaviour of the system with respect to the past in no way contradicts the reversibility of the laws of molecular processes. It is associated with the asymmetry of the conditions of the problem; in the past the system underwent actions from without, while in the future it is left to itself. A more important, but less obvious case is that when the system underwent no action from without but came into a given state (the initial one in our previous consideration) spontaneously, always remaining closed. The question arises as to what was the state occupied by the system before it came into the given state. It could come into a given non-equilibrium state from an even more non-equilibrium state or, on the contrary, from an equilibrium state. But every macroscopic system spends the major part of its

* It should be noted that the drawings presented here and below should be considered as a scheme illustrating the properties of the entropy. In reality, in a system consisting of parts the value of the entropy is strictly defined in the course of a certain finite time interval, but not at every given instant. Hence the graph $S(t)$ should not be taken literally [see B.G.Levich, *Vvedenie v statisticheskuyu fiziku (Introduction to statistical physics)* (Gostekhizdat, Moscow, 1954) §35; F.Reif, *Statistical and thermal physics* (McGraw-Hill, New York, 1965) §15.18].

time in a state of statistical equilibrium. If we ask what state the system was in at $t < 0$, then from the most general considerations it is clear that with a high probability it was in an equilibrium state. Hence a system comes most often into a given non-equilibrium state from an equilibrium state. In other words, in order that a system may come into a given non-equilibrium state at $t = 0$, it must undergo a fluctuation at $t < 0$. In fig. III.15 the solid line represents the most probable process bringing the system into the state which was the initial one for the process presented in fig. III.14. Of course, it is not excluded that the non-equilibrium state given at $t = 0$ arose from another, even more non-equilibrium state, as is shown by the dotted line in fig. III.15. But, since the probability of finding a closed system in a non-equilibrium state is low, such a case is improbable. Now we match the two drawings, i.e. consider the whole process in time. Then we obtain the curves shown in fig. III.16, which are completely symmetric with respect to the future and past.

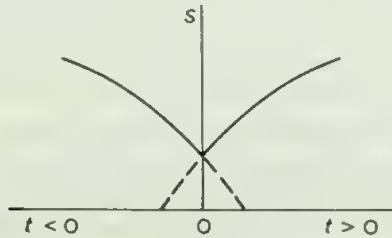


Fig. III.16

The asymmetry of the second law — the indication that the entropy will increase in the future — thus turns out to be associated with the asymmetry of the initial condition — the definition of the system in a non-equilibrium state at the initial instant.

We now imagine that the state of a closed system is defined as the equilibrium state, corresponding to the maximum value of the entropy (fig. III.17). According to the propositions of thermodynamics the system will remain in an equilibrium state, during the subsequent time, and its entropy will remain constant. Statistical physics allows the possibility of a spontaneous transition, a fluctuation, of the system from the equilibrium state. As we have seen above, the probability of a fluctuation decreases sharply with its magnitude. Therefore, in order that we may notice the fluctuation, it is necessary to observe the system for a sufficiently long time interval, in any case much longer than the relaxation time τ . Moreover, the probability of a fluctuation depends very much on the size of the system (the number of particles in it).

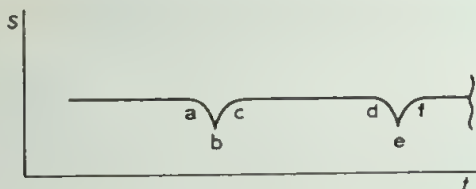


Fig. III.17

If we observe the behaviour of a closed system which is in an equilibrium state at the instant $t = 0$, then we shall see that the entropy of this system will decrease (the segments ab and de on the curve of fig. III.17) and increase (the segments bc and ef) equally frequently. This agrees, after all, with the aforesaid; the process shown in fig. III.16 is, in essence, a particular case of the process shown in fig. III.17, and corresponds to one of the cusps in the latter.

Thus, we see that if one renounces asymmetry in the statement of the problem, i.e. if an improbable (non-equilibrium) state of the system is not given at the initial instant, then the law of increase of entropy loses its one-sided meaning and becomes symmetric with respect to the future and past. This can also be formulated in the following way: in the course of a sufficiently long time interval the number of transitions from an equilibrium state into a non-equilibrium one in a closed system is equal to the number of reverse transitions from a non-equilibrium state into an equilibrium one. This equality arises because the number of the former is equal to a large number of initial (equilibrium) states multiplied by a low probability of the fluctuation. The number of the latter transitions is equal to a small number of initial (non-equilibrium) states multiplied by a high probability of the transition into an equilibrium state (relaxation). In a system which is always closed the entropy increases and decreases equally frequently.

In practice, however, one most often has to deal with systems which are, at the initial instant, in a given non-equilibrium state. In the case of a system which is always closed every non-equilibrium state can be considered as a fluctuation. If we observe the successive changes of state of a system during the course of a time comparable with or smaller than the relaxation time, then the most probable trend of the processes will be an increase of the entropy. In this case the apparent asymmetry arises because of the asymmetry in the statement of the problem.

It is also possible that the system came into a non-equilibrium state as a result of an external action, which thereupon ceases. In this case the system

is insulated only from a certain instant. Its entropy will henceforward increase. Here the asymmetric trend of the entropy is associated with the presence in the past of an action on the system from outside.

We see that the difference between irreversible and reversible processes becomes relative, and that there is no contradiction between the reversibility of the laws of mechanics and the existence of irreversible processes in statistics. In view of the relative character of the ideas of reversibility and irreversibility the necessity arises for a more clear criterion of irreversible and reversible molecular processes. In order to arrive at such a formulation, we shall consider one more concrete example.

Imagine a certain volume in a container occupied by a mixture of two gases. Let, the following non-equilibrium state of the gas be given at a certain initial instant: in the volume considered the composition of the gas deviates by 1% from a homogeneous composition. If the gas is left to itself, then with a probability of the order of unity after the lapse of the relaxation time the molecules will be mixed up and the system will pass over into a state with a uniform density. The entropy of the gas will increase. From the point of view of pure thermodynamics we have a classical example of an irreversible process. However, let us analyse this process more carefully from the statistical point of view. The system will indeed pass over spontaneously from the non-uniform to a uniform distribution of molecules in the mixture. But it cannot be stated that the initial state will never again repeat itself, and that the system will always remain in a state with a uniform distribution of molecules. On the contrary, fluctuations will occur in the system, as a result of which the homogeneity of the gas composition will be violated. After the lapse of a sufficiently long time, in the system left to itself, there will necessarily occur a fluctuation of such a magnitude that the deviation from uniformity in the volume singled out will reach 1%, and the system will come back into the initial state. The return of the system into the initial state shows that the process of mutual diffusion of gases cannot be considered as irreversible. The process considered can be called a reversible process.

One would think that we arrive again at a complete contradiction with pure thermodynamics. However, the practical importance of this contradiction depends on the scale of the phenomenon and the time needed for the system to come back into the initial state. This time can be assumed, roughly speaking, to be inversely proportional to the probability of a fluctuation of the corresponding magnitude. Hence it can be assumed that the time needed for the system to come back into the initial state is longer, the larger the size of the system and the greater the difference between the initial and equilib-

rium state. We shall denote this time, usually called the recovery time, by τ^* . Then it is obvious that, if the total observation time T is small in comparison with τ^* , the system will not manage to come back into the initial state during the observation time. In this case the process accompanied by an increase in entropy is irreversible. But if, on the contrary, the total observation time T is large in comparison with the recovery time τ^* , then during the observation time the system will without fail come back into the initial state. In this case the same process must be considered as reversible. Such a treatment was first developed by M. Smolukhovskii.

Thus, the relation between the times τ^* and T is a decisive factor in the criterion of reversibility and irreversibility.

To get an idea of the order of magnitude of the recovery time τ^* we shall calculate it for a simple system. At the instant $t = 0$ let an ideal gas, confined initially in the left-hand half of a container, occupy the entire container. We shall find the time needed for all N molecules of the gas to gather again, in the left-hand half of the container as a result of molecular motion with a probability close to unity, for example, equal to 0.9.

The probability, for one measurement, that one of the molecules will be found in the left-hand half of the container is equal to $w_1^{(1)} = \frac{1}{2}$. Correspondingly, the probability, for one measurement, that two molecules will be found in the left-hand half of the container is equal to $w_1^{(2)} = (\frac{1}{2})^2$, while the probability of finding in one measurement N molecules in the left-hand half of the container is $w_1^{(N)} = (\frac{1}{2})^N = 2^{-N}$.

The probability that one will not find in one measurement N molecules in the left-hand half of the container is obviously equal to $1 - 2^{-N}$. The probability that one will not find in n measurements N molecules in the left-hand half of the container is equal to $(1 - 2^{-N})^n$.

The probability that after n measurements N molecules will be found in the left-hand half of the container is equal to

$$w_n^{(N)} = 1 - (1 - 2^{-N})^n.$$

Assuming, by our condition, that $w_n^{(N)} = 0.9$, we find

$$n \ln(1 - 2^{-N}) = \ln 0.1.$$

Since N is large, $2^{-N} \ll 1$, so that $\ln(1 - 2^{-N}) \approx -2^{-N}$. Then we have $n/2^N = 1$.

If the measurements are carried out every Δt sec, then N molecules will be found with a probability 0.9 to be gathered again in the left-hand half of

the container after the lapse of a time

$$\tau^* \approx n\Delta t \approx 2^N \Delta t.$$

If, for example, the measurements are carried out every $\Delta t = 1$ sec, then $\tau^* = 2^N$ sec. The recovery time increases very rapidly with increasing number N of particles in the gas.

An idea of the numerical values of τ^* is given by the figures in table 2.

Table 2

N	5	10	100	10^5	10^{19}
τ^* , sec	32	1024	10^{32}	2^{10^5}	$2^{10^{19}}$

We see that, if the number of particles in the system is sufficiently small, the time in which the system comes back into the initial state — the recovery time — can be observed in the course of measurable time intervals.

On the contrary, when the number of particles in the system is large, the recovery time becomes immense. A system with a large number of particles cannot be expected to come back into the initial state in really observable time intervals. It should be noted that experiments have confirmed the numerical values of recovery times for the gathering of a small number of colloidal particles in small volumes (see §56).

From the above it is clear that the point of view developed regarding reversibility and irreversibility does not in practice contradict the conclusions of thermodynamics at all. Recovery times for processes on the macroscopic scale turn out to be so large that for any practical observation time the inequality $\tau^* \gg T$ is always satisfied. Hence processes which are irreversible from the thermodynamic point of view can be considered as irreversible also from the statistical point of view.

We now can pass on to the discussion of the second question put at the beginning of this section: can a perpetual motion machine of the second kind be realized by making use of the phenomenon of fluctuations?

Imagine that we have a certain mechanism which can be used to produce useful work from a fluctuation occurring in a certain system. For concreteness we imagine that this mechanism is a piston which is set in a one-way motion by the density fluctuations occurring in the volume of the gas under the piston. If such a mechanism could be realized in practice, then useful work could be obtained systematically at the expense of the thermal energy of the

medium, i.e. a perpetual motion machine of the second kind could be constructed. However, it is easily shown that the construction of such a machine is impossible.

As a matter of fact, whatever the design of the machine, the piston and the other parts as well as the gas or another medium consist of atoms or molecules. Hence the operating mechanism, as well as the medium, will undergo fluctuations. The fluctuations of the mechanism and the medium are independent of each other and occur, generally speaking, at different instants and in different directions. Let, for example, the piston move and do work as the gas expands. But the piston itself also undergoes fluctuations and hence displacements in the direction opposite to that in which it moves as the gas expands. Owing to the independence of fluctuations in the gas and in the mechanism the displacement of the piston averaged over time turns out to be exactly equal to zero. Consequently, the mean work produced by the piston is also equal to zero.

These qualitative considerations are confirmed by the quantitative calculations for different schemes of such operating mechanisms.

Thus, a systematic production of useful work at the expense of small fluctuations occurring in a certain operating mechanism turns out to be in principle impossible. Similarly, it is impossible to obtain useful work at the expense of single large fluctuations: the probability of large fluctuations decreases incomparably more rapidly than the value of the useful effect increases.

This very fact proves that the construction of a perpetual motion machine of the second kind producing useful work systematically at the expense of fluctuations is in principle impossible. The classical formulation of the second law: "it is impossible to construct a perpetual motion machine of the second kind, i.e. a device which, in the course of a long time, would use heat at a lower temperature and thus serve as a source of a useful work" remains valid.

In conclusion we must dwell on the problem of the so-called "thermal death" of the universe.

The theory of the thermal death of the universe put forward by Clausius consists of the following: For the present the universe is not in a state of thermal equilibrium; there exist in it temperature differences, motion, and so on. However, since the universe represents a closed system to which the laws of statistics and thermodynamics are applicable, after the lapse of a sufficiently long time interval all temperature differences existing in the universe will be smoothed out, and the motion will cease. The universe will pass over into a state of total rest — thermal death.

The teaching of Clausius was subjected to criticism by many physicists, in the first place Boltzmann.

At present it is ascertained without doubt that there are no grounds for applying the laws of statistical physics to the universe changing in time. The treatment of gravitational phenomena within the framework of the general theory of relativity already shows that the thermodynamic properties of systems on a cosmic scale must differ radically from those of ordinary closed systems. In thermodynamics based on the general theory of relativity it turns out that the entropy of systems on a cosmic scale cannot tend to and reach a maximum value, and that thermal equilibrium cannot be established in them *.

A more complete study of the laws of the behaviour of the universe and, in particular, its thermodynamic behaviour, is a matter for the future. But even now it is clear that they are far more complex than the properties of ordinary macroscopic molecular systems, and that the application of the laws of ordinary thermodynamics to the universe is inadmissible.

* See R. Tolman, *Relativity, thermodynamics and cosmology* (Oxford University Press, Oxford, 1934).

Ideal Gases

§37. The distribution function for ideal gases

In this and following chapters we shall consider the application of the general theory to concrete systems. In the first place ideal gases will be considered.

The partition function of an ideal monatomic gas has the form

$$Z = \sum e^{-\epsilon_n/kT} \Omega(\epsilon_n), \quad (37.1)$$

where ϵ_n is the energy of the gas as a whole, and the summation is carried out over the energy levels of the system.

Since the spacings between the energy levels of the gas as a whole are very small in comparison with kT , the summation over the energy levels can be replaced by integration. Thus,

$$Z = \int e^{-\epsilon/kT} d\Omega. \quad (37.2)$$

Let us find the number of states, $d\Omega$, of a system with given energy. Since all particles of an ideal gas are independent, one can write that

$$d\Omega = \prod_i' d\Omega_i = \prod_i' \frac{dp_i dq_i}{h^{3N}},$$

where the product is taken over all coordinates and momenta of the particles. The prime in the product denotes that in its formation one has to include only those terms which correspond to different states of the system as a whole.

Substituting $d\Omega$ into (37.1) we find

$$Z = \frac{\int e^{-\epsilon/kT} \prod_i' dp_i dq_i}{h^{3N}}. \quad (37.3)$$

In order to find the integration range in (37.3), we shall consider the case of two uniformly moving particles. Let p_1 denote the momentum of the first particle and p_2 the momentum of the second particle. The integration in (37.3) is carried out over all values of the momenta which each of the molecules can have. We shall consider the two states of the system shown in fig. III.18 by dotted lines. In the first state the first molecule has a momen-

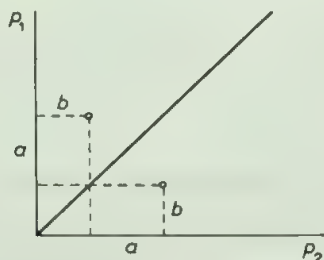


Fig. III.18

tum equal to a , and the second molecule has a momentum equal to b . In the second state, on the contrary, the second molecule has momentum a , and the first molecule has momentum b . It can be said that the second state differs from the first by the fact that the momentum of the first particle is replaced by that of the second particle, and vice versa. In other words, the representative points of the two molecules, or for brevity, the molecules, are mutually exchanged in phase space.

In carrying out the integration in (37.3) we take into account separately each of these and similar states of the system of two molecules. Integrating with respect to p_1 at a fixed value of p_2 , and thereupon with respect to p_2 at a fixed value of p_1 , we pass through a sequence of values of the momenta (p_1, p_2) and (p_2, p_1) , differing from each other by the mutual exchange of the particles in phase space. In other words, we assume that the states "the first molecule has a momentum p_1 , and the second molecule has a momentum p_2 " and "the first molecule has a momentum p_2 and the second molecule has a momentum p_1 " are two different states. However, if the two molecules are identical, then these states are equivalent. But from the proposition of quantum theory on the complete identity of elementary particles it follows that the two states correspond to one and the same physical state of the system. A physical state of the system is characterized by the fact that one particle has a momentum p_1 , and the second particle has a momentum p_2 . It does not matter which of the particles has the momentum p_1 or p_2 , since the two particles are completely identical. The states (p_1, p_2) and (p_2, p_1) are not two equivalent states, but one and the same state of the system. Hence we would make a mistake by taking the two states (p_1, p_2) and (p_2, p_1) as independent states in the integration in (37.3). In reality both of these correspond to only one physical state of the system. In order to avoid this mistake one should take into account only one of the states: either (p_1, p_2) or (p_2, p_1) . The integration with respect to p_1 and p_2 in (37.3) should be carried out not over all their possible values but only over the values corresponding to physically different states of the system. For this one can, for example, integrate not over the entire phase area (p_1, p_2) but over one half of it, cut by the bisector drawn in fig. III.18. However, it is simpler to proceed differently. One can, as before, carry out the integration over all values of p_1 and p_2 , and reduce by a factor of two the result obtained. This will compensate for the incorrect doubling of the number of states as they are calculated. One has to proceed completely analogously also in calculating that part of the partition function which includes the integration with respect to the coordinates of the two molecules. States differing from each other only by the mutual exchange of the molecules in space should be considered not as different states, but as one and the same state.

In general, in calculating the integral over all states of the system consisting of two particles one has to carry out the integration over all states — the coordinates and momenta (p_1, q_1) of the first particle and (p_2, q_2) of the second particle, but the result obtained should be reduced by a factor of two. In this case it will be automatically taken into account that the "exchange" in phase space of the representative points characterizing the state of each of the particles does not lead to different states of the entire system.

The result obtained can be generalized to arbitrary states of a gas containing N molecules. All states of the gas which differ from each other only by the fact that the values of the coordinates and momenta of one molecule are exchanged with those of another molecule are identical physical states. It can be said that the set of all states which are obtained by the mutual permutation of N representative points in the phase space corresponds to only one physical state. In calculating the total partition function of the gas each of these must be taken into account in the integration only once. Hence in carrying out the integration over all possible values of the coordinates and momenta of gas molecules the result must be divided by the number of the mutual permutations of representative points in the phase space. This number is, obviously, equal to $N!$.

It should be noted that the necessity of dividing the phase integral by $N!$ was taken into account in classical statistics. Otherwise the additivity of the thermodynamic functions obtained would be violated. However, the full meaning of this procedure is seen in quantum statistics after taking into consideration the principle of the identity of atomic particles.

Thus, we can write that

$$d\Omega = \prod_i' d\Omega_i = \frac{\prod_i d\Omega_i}{N!} = \frac{1}{N!} \prod_i \frac{dp_i dq_i}{h^{3N}}. \quad (37.4)$$

The Gibbs distribution for an ideal gas has the form

$$dw = \frac{1}{Z} e^{-\epsilon/kT} d\Omega = \frac{1}{Z} e^{-\epsilon/kT} \frac{dp_1 \dots dp_{3N} dq_1 \dots dq_{3N}}{h^{3N} N!}.$$

Correspondingly, for the partition function we obtain

$$Z = \frac{1}{N!} \int e^{-\epsilon/kT} \prod_i d\Omega_i,$$

where the integration is carried out over the entire phase space. Writing the energy of the gas as follows:

$$\epsilon = \sum_i \epsilon_i,$$

we obtain

$$Z = \frac{1}{N!} \frac{1}{h^{3N}} \prod_i \int e^{-\epsilon_i/kT} dp_i dq_i = \frac{z^N}{N!},$$

where z is the partition integral for one molecule calculated earlier [formula (19.7)]. Substituting its value, we find that

$$Z = \frac{1}{N!} \left(\frac{2\pi mkT}{h^2} \right)^{\frac{3}{2}N} V^N. \quad (37.5)$$

We pass on to the calculation of the thermodynamic functions of an ideal monatomic gas.

For the energy of the gas one can write

$$E = kT^2 \frac{\partial}{\partial T} \ln Z = \frac{3}{2} NkT. \quad (37.6)$$

The energy of the gas is proportional to the temperature and does not depend on the volume of the gas, in correspondence with formula (20.14).

The heat capacity of a monatomic gas turns out to be equal to

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = \frac{3}{2} Nk \approx 12 \text{ J mol}^{-1} \text{ K}^{-1}.$$

For the free energy of an ideal monatomic gas we find, according to (32.1),

$$F = -kT \ln Z = -NkT \ln V \left(\frac{2\pi mkT}{h^2} \right)^{\frac{3}{2}} + kT \ln N!$$

In order to calculate $N!$ for a large N one can make use of Stirling's formula (see Appendix IV). We then have

$$F = -NkT \ln \left[\frac{eV}{N} \left(\frac{2\pi mkT}{h^2} \right)^{\frac{3}{2}} \right]. \quad (37.7)$$

From the free energy one can find the equation of state of the gas. In correspondence with (32.6) we obtain

$$p = - \left(\frac{\partial F}{\partial V} \right)_T = \frac{NkT}{V}. \quad (37.8)$$

Thus, we arrive at the well-known equation of state of an ideal gas. It should be noted that we have found this equation in a purely theoretical way, without any reference to experimental data. Only the numerical value of the constant k is determined experimentally.

We shall now calculate the entropy of the gas. According to (29.7) it is equal to

$$\begin{aligned} S &= - \left(\frac{\partial F}{\partial T} \right)_V = Nk \ln \left[\frac{eV}{N} \left(\frac{2\pi mkT}{h^2} \right)^{\frac{3}{2}} \right] + \frac{3}{2}Nk = \\ &= Nk \ln(V/N) + C_V \ln kT + \frac{5}{2}Nk + Nkj. \end{aligned} \quad (37.9)$$

For reasons which will be explained later the quantity

$$j = \ln(2\pi m/h^2)^{\frac{3}{2}}$$

is called the chemical constant.

The entropy defined by formula (37.9) is expressed as a function of the number of particles and the volume of the gas.

The entropy is proportional to the number of particles of the gas as it should be, by virtue of its additivity. As the volume and number of particles of the gas are simultaneously increased by an arbitrary factor, the entropy increases or decreases by the same factor.

It may seem strange that the entropy which we have calculated does not contain any arbitrary constant, while, according to what was said in §24, the entropy is determined only to within an indefinite constant. In fact, the formula does not contain any arbitrary constant because we based the calculation of F and S on formulae (29.7) and (32.1), in which the arbitrary constant of the entropy was chosen as the conditional zero entropy.

The expression (37.9) for entropy loses its applicability as $T \rightarrow 0$. In deriving (37.9) we have not taken into account the phenomenon of quantum degeneration of the gas, which plays a basic role in its behaviour at very low temperatures. This phenomenon will be considered in ch. 10.

One cannot obtain a correct expression for the entropy, satisfying the third law of thermodynamics, without taking into account the phenomenon of degeneration. However, it should be noted that there are no ordinary gases at such low temperatures as those at which degeneration occurs. Long

before these temperatures are reached the liquefaction of gases takes place for all practical values of the density.

At very high temperatures formula (37.9) is again inapplicable, since it does not take into account the thermal ionization of atoms. Nevertheless, the interval of applicability of formula (37.9) is very large: from the liquefaction point to several thousand degrees.

In practice it is often more convenient to express entropy in terms of the pressure and temperature. Substituting into (37.9) the value of V expressed in terms of the pressure, we find

$$\begin{aligned} S &= \frac{5}{2}Nk \ln kT - Nk \ln p + \frac{5}{2}Nk + Nkj = \\ &= C_p \ln kT - Nk \ln p + \frac{5}{2}Nk + Nkj. \end{aligned} \quad (37.10)$$

In addition we calculate the Gibbs thermodynamic potential G . On the basis of (28.11) and (37.7), we have for G :

$$\begin{aligned} G = F + pV &= -\frac{5}{2}NkT \ln kT + NkT \ln p - NkTj = \\ &= NkT \ln p - C_p T \ln kT - NkTj. \end{aligned} \quad (37.11)$$

The expression for the heat capacity obtained above can be compared with experimental data. The number of monatomic gases is not very large. Noble gases and metal vapours are examples.

Table 3 gives measured values of the heat capacity.

Table 3
Heat capacity (at a constant volume) of monatomic gases

Substance	Temperature (°C)	C_V (J mol ⁻¹ K ⁻¹)
He {	291	12.59
	93	12.26
	26	12.56
	18	12.64
Ar	288	12.85
Na (vapour)	750–920	12.35
Hg (vapour)	548–629	12.43

From the table it is seen that the predictions of theory are well justified experimentally: the heat capacity of monatomic gases is constant over a large temperature interval and has a value which is almost the same as the theoretical value:

$$C_V \approx 12 \text{ J mol}^{-1} \text{ K}^{-1}.$$

In conclusion we shall consider the important problem of gas mixing.

For simplicity we assume that gas is put into two chambers with volumes V_1 and V_2 , which are separated in the beginning by an impenetrable membrane. Then the membrane is removed, and the molecules of both portions of the gas begin to interdiffuse. As a result of this process gas mixing takes place. We assume that the temperature and pressure of the gases were equal before mixing.

Let us find the change in the entropy as the two portions of the gas mix with each other. Here two cases should be distinguished: the mixing of gases of a different kind, and the mixing of gases of the same kind.

We shall begin with the treatment of the first case.

According to (37.9), the entropies of the different gases before mixing are given by the expressions

$$S_1^{(0)} = N_1 k \ln \frac{V_1}{N_1} + N_1 f(T),$$

$$S_2^{(0)} = N_2 k \ln \frac{V_2}{N_2} + N_2 f(T),$$

where $f(T)$ are the terms of the formula for the entropy which do not depend on the volume. The total entropy of the system before mixing is equal to

$$S^{(0)} = S_1^{(0)} + S_2^{(0)}.$$

After mixing, each of the ideal gases will behave as though there were no second gas, and will occupy the entire volume $V_1 + V_2$. The temperature of the mixture will be equal to the initial temperature of the gas. Hence after the mixing the entropy of each of the gases will be equal to

$$S_1 = N_1 k \ln \frac{V_1 + V_2}{N_1} + N_1 f(T),$$

$$S_2 = N_2 k \ln \frac{V_1 + V_2}{N_2} + N_2 f(T).$$

The entropy of a mixture consisting of two non-interacting ideal gases is equal to the sum of their entropies, i.e.

$$S = S_1 + S_2.$$

The change in the total entropy of the entire system in mixing is equal to $\Delta S = S - S^{(0)} =$

$$\begin{aligned} N_1 k \ln \frac{V_1 + V_2}{N_1} + N_2 k \ln \frac{V_1 + V_2}{N_2} - N_1 k \ln \frac{V_1}{N_1} - N_2 k \ln \frac{V_2}{N_2} = \\ = N_1 k \ln \frac{V_1 + V_2}{V_1} + N_2 k \ln \frac{V_1 + V_2}{V_2}. \end{aligned}$$

At given temperature and pressure

$$\frac{V_1 + V_2}{V_1} = \frac{(N_1 + N_2)kT/p}{N_1 kT/p} = \frac{N_1 + N_2}{N_1},$$

and analogously for $(V_1 + V_2)/V_2$, so that the change in the entropy is equal to

$$\Delta S = N_1 k \ln \frac{N_1 + N_2}{N_1} + N_2 k \ln \frac{N_1 + N_2}{N_2}.$$

Thus, the entropy of the mixture is higher than that of the initial gases.

The process of mixing of two different gases is an irreversible process. The origin of this irreversibility is quite understandable. When the membrane separating the gases is removed the interdiffusion of the gases takes place. Before mixing there existed a "regularity" in the position of molecules: the molecules of one gas were in one part of the container, and those of the second gas were in the other part of the container.

When the diffusion completely mixed the two gases, a uniform completely random distribution of the molecules occurs, and the probability of the state increases. In order to separate the gases it is necessary to do a certain amount of work, which in principle can be calculated.

Now consider the process of mixing two identical gases. Gases can be considered as identical in the case where they behave identically in all possible external fields.

The entropy of two portions of one gas before mixing is equal to

$$S_1 = N_1 k \ln \frac{V_1}{N_1} + N_2 k \ln \frac{V_2}{N_2} + (N_1 + N_2) f(T).$$

The entropy of the entire gas after mixing is equal to

$$S_2 = (N_1 + N_2) k \ln \frac{V_1 + V_2}{N_1 + N_2} + (N_1 + N_2) f(T).$$

Then for the change in the entropy we obtain

$$\Delta S = S_2 - S_1 = (N_1 + N_2) k \ln \frac{V_1 + V_2}{N_1 + N_2} - N_1 k \ln \frac{V_1}{N_1} - N_2 k \ln \frac{V_2}{N_2}.$$

However, from the equation of state of the gas it follows that, at a constant pressure and temperature,

$$\frac{V_1 + V_2}{N_1 + N_2} = \frac{V_1}{N_1} = \frac{V_2}{N_2},$$

Hence

$$\Delta S \equiv 0.$$

Thus, the change in the entropy in mixing two portions of one gas is indeed identically equal to zero. This result, which is in complete agreement with experiment, is closely connected with the assumption of the mutual identity of all particles of the given gas. Owing to this identity their mutual mixing is not a physical event. In mixing two portions of one gas at a constant pressure and temperature the distribution of molecules in the overall volume turns out to be uniform and random, and no interdiffusion takes place. It should be stressed that molecules or atoms can be considered as belonging to one kind and, hence, as identical, only in the case where they have the identical chemical structure, mass, and all other characteristics. This means that even different isotopes of one element, or atoms in different energy states, cannot be considered as identical. Thus, for example, the mixing of two different isotopes of a gas, or two portions of a gas consisting of normal and excited molecules, represents an irreversible process. This is particularly clear from the fact that a spontaneous separation of mixed gases does not occur, and that a certain amount of work must be done for their separation.

§38. The Maxwell–Boltzmann distribution and the Boltzmann distribution in a uniform field of force

In practice one often has to deal with a gas in a uniform external field of force. The most important example of such a field is the gravitational field. Up to now we have not considered the action of the gravitational field on the behaviour of a gas. We shall now consider an ideal gas in a uniform field of force. In such a field every molecule has a total energy

$$\epsilon = \epsilon_{\text{transl}} + u(x, y, z),$$

where ϵ_{transl} is the kinetic energy of its translational motion, and u is the potential energy in the external field, depending on the position of the particle.

Substituting this expression for the energy into the Gibbs distribution (19.6) for a molecule of an ideal gas, we have

$$dw = \frac{1}{zh^3} \exp\left[\frac{\epsilon_{\text{transl}} + u}{kT}\right] dp_x dp_y dp_z dV, \quad (38.1)$$

where the integral over states is, obviously, equal to

$$z = \int \exp\left[\frac{\epsilon_{\text{transl}} + u}{kT}\right] \frac{dp_x dp_y dp_z dV}{h^3}. \quad (38.2)$$

The integration is carried out over all possible values of the variables. Noting that the integral over states can be written in the form

$$\begin{aligned} z &= \int e^{-\epsilon_{\text{transl}}/kT} \frac{dp_x dp_y dp_z}{h^3} \int e^{-u/kT} dV = \\ &= \left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}} \int e^{-u/kT} dV, \end{aligned} \quad (38.3)$$

we find that the Gibbs distribution, normalized to unity, for a molecule of an ideal gas in the presence of an external field has the form

$$dw = \left[\frac{1}{(2\pi mkT)^{\frac{3}{2}}} e^{-p^2/2mkT} dp_x dp_y dp_z \right] \left[\frac{e^{-u/kT} dV}{\int e^{-u/kT} dV} \right]. \quad (38.4)$$

The probability distribution obtained, characterizing the probability for a molecule to have a given momentum and to be in a given volume element is called the Maxwell-Boltzmann distribution.

The first of the factors in (38.4) is the Maxwell distribution, with which we are familiar. It characterizes the probability distribution over the components of the momentum. The second factor depends only on the coordinates of the molecule and is determined by the form of its potential energy $u(x, y, z)$ in an external field of force. It expresses the probability for a molecule to be found in a given volume dV . In the particular case where there is no external field the distribution of molecules over the entire volume of the container is uniform and the second factor reduces to the value $V^{-1}dV$.

On the basis of the theorem of multiplication of probabilities the Maxwell-Boltzmann distribution can be considered as the product of the probabilities of two independent events: the probability of a given value of the momentum and the probability of a given position of the molecule. The first probability represents the Maxwell distribution, while the second probability represents the Boltzmann distribution. Each of the distribution is normalized to unity.

The fact that the two distributions are independent expresses an important, and at first not obvious physical proposition: the probability of a given value of momentum does not depend on the position of the molecule and, conversely, the probability of a given position of the molecule does not depend on its momentum.

Now consider in more detail the Boltzmann distribution for the particular case where the gas is in the field of terrestrial gravitation. We direct the z -axis vertically up. Then the potential energy of a gas molecule can be written in the form

$$u = mgz .$$

Since the potential energy depends only on the height, molecules are distributed uniformly in the plane $z = \text{const}$. Hence only the dependence of the probability distribution on the coordinate z is of interest. It has the form

$$dw_B = \frac{e^{-mgz/kT} dz}{\int e^{-mgz/kT} dz} , \quad (38.5)$$

where the integral is taken over all possible values of z .

Introducing the mean number of particles per cm^3 at a given height instead of the probability distribution, one can rewrite (38.5) in the form

$$dn = n_0 e^{-mgz/kT} dz, \quad (38.6)$$

where n_0 is the number of particles per cm^3 in the plane $z = 0$.

Formula (38.6) shows that the density of a gas in the gravitational field decreases according to an exponential law. It decreases by a factor e as the height increases to $\delta = kT/mg$. This quantity can be called the characteristic length of the distribution of particles in the gravitational field. For hydrogen δ amounts to about 3×10^5 m at room temperature while for air δ is correspondingly equal to 10^4 m. The Maxwell velocity distribution of molecules with a constant temperature T holds at all heights. However, the number of molecules at different heights decreases according to the exponential law (38.6). At first sight the constancy of the temperature at all heights may seem to contradict the following simple reasoning: if a molecule, which at a height z_0 has a kinetic energy $\frac{1}{2}mv_0^2$ moves to a height z , then its kinetic energy must decrease to a value $\frac{1}{2}mv_0^2 - mg(z - z_0)$, where $mg(z - z_0)$ is the work done against the force of gravity. Hence at a great height a molecule will have smaller velocity and kinetic energy. But, on the other hand, the temperature is connected with the mean square velocity by the relation (10.6). Consequently, the temperature of the gas must decrease with height. The fallacy of this reasoning lies in the consideration of only one molecule, without taking into account its collisions with other gas molecules. The Maxwell velocity distribution is established owing to collisions between molecules. In the foregoing reasoning the establishment of the Maxwell distribution is ignored and the "temperature of a molecule", which has no meaning, is considered. In fact, those molecules which have a large velocity will preferentially go up. Hence the Maxwell distribution will be established automatically at all heights.

Let us consider certain conclusions which can be derived from the density distribution of the gas with height. First of all, we shall dwell on the idea of the weight of a gas. Imagine a container of a height h , confining a gas. This gas has a certain weight. It is often said that the weight of a gas is the weight of all the molecules constituting it. In reality, however, this is not strictly true. The weight of a gas is measured by the difference between the pressures exerted by the gas on the bottom and the top of the container. All molecules of the gas, which are in permanent motion and which in the greater part of the container do not collide directly either with the bottom or the top of the container, take part in the production of this pressure difference. In this lies the difference between the weight of a gas and the weight of a body lying on the pan of a balance.

Let us find the weight of a gas column of height h . We can proceed in two

ways. First, it can be determined purely formally, by writing that the weight of the gas column is equal to the weight of all the molecules constituting it. Second, it can be found by taking the difference between the pressures exerted by the gas on the bottom ($z=0$) and the top ($z=h$) of the container.

In the first case we have:

$$\begin{aligned} P &= mg \int dN = mgn_0 \int_0^h e^{-mgz/kT} dz \int dx dy = \\ &= Smgn_0 \frac{kT}{mg} (1 - e^{-mgh/kT}) = SkT(n_0 - n_h), \end{aligned}$$

where S is the area of the cross section of the container, and n_0 and n_h are the densities of the gas at the heights $z = 0$ and $z = h$.

In the second case we can write that

$$P = S(p_0 - p_h) = S(n_0 - n_h)kT,$$

where p_0 and p_h are the pressures at the heights $z = 0$ and $z = h$ respectively. Thus, the calculation confirms the validity of the concept of the weight of a gas as measured by the difference between the pressures on the bottom and top of the container.

The molecules of a gas which are in a gravitational field possess a certain mean potential energy which exceeds the mean energy of the gas outside the field of force. Hence the mean energy of a gas in a gravitational field and, consequently, also its heat capacity must be higher than the values which we have calculated earlier. Let us find the extra heat capacity of the gas in a gravitational field. For this we shall calculate the mean potential energy of a gas molecule in a gravitational field. By definition it is equal to

$$\bar{u} = mg\bar{z} = mg \int z dw_B, \quad (38.7)$$

where dw_B is the probability that the molecule will be found at a height between z and $z + dz$, given by formula (38.5). Substituting the expression for dw_B into (38.7), we have

$$\bar{u} = mg\bar{z} = mg \frac{\int z e^{-mhz/kT} dz}{\int e^{-mgz/kT} dz}. \quad (38.8)$$

In calculating the integrals contained in (38.8) it is essential to know the

height of the gas column. Consider first of all the case of an infinitely high gas column or, more precisely, a column confined in a container whose height is considerably larger than the characteristic height δ . Then the range of integration with respect to z extends from $z = 0$ to $z \rightarrow \infty$. The calculation of the simple integrals (see Appendix IV) gives

$$\bar{u} = kT.$$

The mean potential energy of one molecule in the infinitely high gas column turns out to be proportional to the absolute temperature. The mean potential energy of a gram-mole of the gas in the infinitely high column is equal to

$$\bar{U} = N_0 \bar{u} = N_0 kT.$$

Whence we find for the extra heat capacity at a constant volume per gram-mole due to the potential energy

$$C_V^{\text{pot}} = N_0 k.$$

The heat capacity obtained is thus comparable with that due to the kinetic energy of the molecules.

We come to a completely different conclusion in the case of a gas column confined in a container of a height $h \ll \delta$. In this case the mean potential energy of the molecules is equal to

$$\bar{u} = mg \frac{\int_0^h e^{-mgz/kT} z dz}{\int_0^h e^{-mgz/kT} dz}.$$

Since in the integration range $mgz/kT = z/\delta \ll 1$, one can expand the integrand in a series and confine oneself to the first term of the expansion. This gives

$$\bar{u} = \frac{1}{2} mgh, \quad \bar{U} = N\bar{u} = \frac{1}{2} Nmgh.$$

In this approximation the mean potential energy of the gas column does not depend on the temperature at all. The corresponding contribution to the heat capacity is zero. In the next higher order approximation one can obtain a heat capacity which is very small in comparison with that due to the presence of kinetic energy. Thus, in practical cases of interest the contribution of the potential energy to the heat capacity of a gas can be disregarded.

Finally, consider the height distribution of molecules possessing different masses. From the form of the distribution (38.5) it is clear that the larger the mass of the molecule the more rapidly the number of corresponding molecules decreases with height. If there are equal numbers of molecules with masses m_1 and m_2 at the height $z = 0$, then the ratio of the numbers of the molecules of the two kinds at a height h is equal to

$$\frac{n_1}{n_2} = \exp \left[- \frac{(m_1 - m_2)gh}{kT} \right].$$

If the simple laws of the equilibrium distribution of density were applicable to the Earth atmosphere, then a sharp change of the composition of the atmosphere with height would be observed. In reality, however, measurements performed on the composition do not confirm this conclusion. It is also well known that the temperature decreases with height, which is also in complete contradiction with the requirement of the constancy of the temperature in an equilibrium gas column. These, and also a number of other facts show that the atmosphere is not in a state of statistical equilibrium.

§39. The calculation of the heat capacity of diatomic molecules by means of classical statistics. The law of equipartition of energy over degrees of freedom

Most substances in the gaseous state exist in the form of molecules. One very often has to deal with diatomic gases. Examples of such gases are H_2 , O_2 , N_2 , HCl and CO . Our immediate problem is the generalization of the results previously obtained to the cases of polyatomic gases and, in particular, to diatomic gases.

The basic difference between diatomic and polyatomic gases and monatomic gases is the presence of rotational and vibrational degrees of freedom in the former. We shall begin the treatment with the simplest case, diatomic gases. We shall assume at first that the molecule represents a system obeying the laws of classical mechanics. Since we shall not be interested in the internal motion of electrons in the atom, we shall replace each atom by a material point having no extension. Two material points bound into a molecule can be likened to a miniature dumb-bell at the ends of which there are two infinitely small spheres with masses m_1 and m_2 (different atomic masses). Since the link between the atoms in the molecule is not absolutely rigid, the following forms of motion are possible: translational motion (three degrees of freedom),

rotation about two axes perpendicular to the axis connecting the two atoms (two degrees of freedom), and vibration of the atoms along the line connecting them. We consider molecules as material points having infinitesimal dimensions, so that it makes no sense to speak of the rotation about the axis of the molecule.

The energy of the molecule can be written in the form

$$\epsilon = \epsilon_{\text{transl}} + \epsilon_{\text{rot}} + \epsilon_{\text{vib}}, \quad (39.1)$$

where ϵ_{transl} , ϵ_{rot} and ϵ_{vib} are respectively the energies of the translational, rotational and vibrational motion.

The state of a molecule formed by two atoms bound to each other is determined by defining six coordinates and six momenta. The coordinates are: the x, y, z coordinates determining the position of the centre of mass in space; θ and φ determining the position of the axis of the molecule in space; the q -coordinate determining the departure of the atoms from the equilibrium distance. The momenta corresponding to these coordinates are

$$p_x, \quad p_y, \quad p_z, \quad M_1 = I\omega_1, \quad M_2 = I\omega_2, \quad p_q.$$

Here I denotes the moment of inertia, and ω denotes the angular velocity of rotation of the molecule. The corresponding phase space has $2 \cdot (3+2) = 12$ dimensions.

We apply the Gibbs distribution to the molecule, considering it as a quasi-closed subsystem. Obviously, we have

$$dw = \frac{1}{h^{6z}} \exp \left[-\frac{\epsilon_{\text{transl}} + \epsilon_{\text{rot}} + \epsilon_{\text{vib}}}{kT} \right] d\Gamma. \quad (39.2)$$

An element of the phase space of a diatomic molecule can be written in the form

$$d\Gamma = d\Gamma_{\text{transl}} d\Gamma_{\text{rot}} d\Gamma_{\text{vib}}, \quad (39.3)$$

where the following notations are introduced:

$$\begin{aligned} d\Gamma_{\text{transl}} &= dp_x dp_y dp_z dx dy dz; \\ d\Gamma_{\text{rot}} &= dM_1 dM_2 \sin \theta d\theta d\varphi; \\ d\Gamma_{\text{vib}} &= dp_q dq. \end{aligned} \quad (39.4)$$

The first factor in the expression of the phase volume corresponds to the three degrees of freedom of the translational motion, the second factor corresponds to the two degrees of freedom of the rotational motion, and the third factor corresponds to the vibrational motion of the molecule. Thus,

$$dw = \frac{1}{h^6} \exp \left[-\frac{\epsilon_{\text{transl}} + \epsilon_{\text{rot}} + \epsilon_{\text{vib}}}{kT} \right] d\Gamma_{\text{transl}} d\Gamma_{\text{rot}} d\Gamma_{\text{vib}}. \quad (39.5)$$

We see that the Gibbs distribution resolves into three independent factors, corresponding to the translational, rotational and vibrational motions. Each of these forms of motion is independent of the other two. Hence the translational motion, rotation and vibration can be considered independently of each other.

The translational motion of diatomic molecules differs in no way from that of monatomic molecules, since the translational motion reduces to the motion of the centre of mass of the system.

Now let us consider *the rotational motion of a diatomic molecule*.

The rotational energy of a diatomic molecule for a given distance between the atoms has the form

$$\epsilon_{\text{rot}} = \frac{I\omega_1^2}{2} + \frac{I\omega_2^2}{2} = \frac{M_1^2}{2I} + \frac{M_2^2}{2I}, \quad (39.6)$$

where ω_1 and ω_2 are the angular velocities of the molecule; and M_1 and M_2 are the angular momenta. The moment of inertia I of the molecule is equal to

$$I = \frac{m_1 m_2}{m_1 + m_2} a^2, \quad (39.7)$$

where a is the distance between the atoms, and m_1 and m_2 are their masses.

The probability for the molecule to have values of angular momenta lying between M_1 and $M_1 + dM_1$, M_2 and $M_2 + dM_2$, and to be oriented in space in such a way that its axis forms with the coordinate axes angles between θ , $\theta + d\theta$ and φ , $\varphi + d\varphi$, has the form

$$dw_{\text{rot}} = \text{const} \cdot \exp \left[-\frac{M_1^2 + M_2^2}{2IkT} \right] dM_1 dM_2 \sin \theta d\theta d\varphi. \quad (39.8)$$

Since we are not interested in the orientation of the molecule in space, it is more convenient to pass over from the expression (39.8) to the expression

for the probability that the angular momentum has a given value for any orientation of the axis of the molecule in space. Integrating expression (39.8) with respect to the angles θ and φ , we obtain

$$dw_{\text{rot}} = \text{const} \cdot \exp \left[-\frac{M_1^2 + M_2^2}{2IkT} \right] dM_1 dM_2. \quad (39.9)$$

The constant in the expression (39.9) can be found from the normalization condition.

Instead of the angular momenta with respect to the axes, one can introduce the more customary quantities, the angular velocities. Then the probability for the molecule to have angular velocity components lying between ω_1 , $\omega_1 + d\omega_1$, ω_2 , $\omega_2 + d\omega_2$ is given by

$$dw_{\text{rot}} = \text{const} \cdot \exp \left[-\frac{I(\omega_1^2 + \omega_2^2)}{2kT} \right] d\omega_1 d\omega_2, \quad (39.10)$$

where the constant is again determined from the normalization condition.

Integrating formula (39.10) over all values of the angular velocity components (the integration range can be extended up to $\pm \infty$ on the same grounds as for the translational velocity components), we have

$$1 = \text{const} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left[-\frac{I(\omega_1^2 + \omega_2^2)}{2kT} \right] d\omega_1 d\omega_2, \quad (39.11)$$

whence

$$\text{const} = \frac{I}{2\pi kT}. \quad (39.12)$$

Hence we have finally

$$dw_{\text{rot}} = \frac{I}{2\pi kT} \exp \left[-\frac{I(\omega_1^2 + \omega_2^2)}{2kT} \right] d\omega_1 d\omega_2. \quad (39.13)$$

Proceeding from the condition (39.10), we find the mean rotational energy

$$\overline{\epsilon}_{\text{rot}} = \frac{1}{2} I (\overline{\omega_1^2} + \overline{\omega_2^2}). \quad (39.14)$$

For the mean value of the square of the angular velocity component we have

$$\overline{\omega_1^2} = \frac{I}{2\pi kT} \int_{-\infty}^{\infty} \omega_1^2 e^{-I\omega_1^2/2kT} d\omega_1 \int_{-\infty}^{\infty} e^{-I\omega_2^2/2kT} d\omega_2 = \frac{kT}{I}. \quad (39.15)$$

From consideration of the symmetry it is clear that

$$\overline{\omega_1^2} = \overline{\omega_2^2}, \quad (39.16)$$

hence the mean value of the rotational energy is equal to

$$\overline{\epsilon}_{\text{rot}} = kT. \quad (39.17)$$

To each degree of freedom of the rotational motion there corresponds an energy equal to $\frac{1}{2}kT$.

Passing on to the *vibrational motion of a diatomic molecule*, we note first of all that in the first approximation one can confine oneself to small vibrations about the equilibrium position, i.e. the equilibrium distance between the two atoms. In this case the vibrational energy of a diatomic molecule can be written in the form

$$\epsilon_{\text{vib}} = \frac{\mu \dot{q}^2}{2} + \frac{\mu \omega^2 q^2}{2} = \frac{p_q^2}{2\mu} + \frac{\mu \omega^2 q^2}{2}, \quad (39.18)$$

where q is the departure of the atoms from the equilibrium position, μ is the reduced mass, and ω is the vibrational frequency connected with the constant of the quasi-elastic force κ by the relation $\omega = (\kappa/\mu)^{\frac{1}{2}}$. The momentum of the vibrating system is $p_q = \mu \dot{q}$.

The probability for the atoms in the vibrating molecule to be in a position q and to have a momentum p_q has the form

$$dw_{\text{vib}} = \frac{1}{z} \exp \left[-\frac{p_q^2}{2\mu kT} \right] \exp \left[-\frac{\mu \omega^2 q^2}{2kT} \right] \frac{dp_q dq}{h}. \quad (39.19)$$

The partition function z in the expression (39.19) is found from the normalization condition:

$$z = \frac{1}{h} \int_{-\infty}^{\infty} \exp \left[-\frac{p_q^2}{2\mu kT} \right] dp_q \int_{-\infty}^{\infty} \exp \left[-\frac{\mu \omega^2 q^2}{2kT} \right] dq. \quad (39.20)$$

Since the integrands decrease very rapidly with increasing arguments and the integrals converge rapidly, we have extended the integration range to infinity.

A simple calculation gives

$$z = \frac{2\pi kT}{h\omega} . \quad (39.21)$$

Thus, we have finally

$$dw_{\text{vib}} = \left(\frac{\omega}{2\pi kT} \right) \exp \left[-\frac{p_q^2}{2\mu kT} \right] \exp \left[-\frac{\mu\omega^2 q^2}{2kT} \right] dp_q dq . \quad (39.22)$$

We find the mean energy of the vibrational motion:

$$\bar{\epsilon}_{\text{vib}} = \frac{\overline{p_q^2}}{2\mu} + \frac{\mu\omega^2 \overline{q^2}}{2} . \quad (39.23)$$

Calculating the means, we have

$$\overline{p_q^2} = \mu kT , \quad \overline{q^2} = \frac{kT}{\mu\omega^2} . \quad (39.24)$$

Substituting the mean values from (39.24) into (39.23), we find

$$\bar{\epsilon}_{\text{vib}} = kT . \quad (39.25)$$

To one vibrational degree of freedom there corresponds on the average an energy two times larger than that corresponding to one degree of freedom of the translational or rotational motion. The meaning of this will become clear if one recalls that for the vibrational motion the mean (over one period) kinetic energy of the system is equal to the mean potential energy. The energy of vibrational motion consists of two components having the same structure: quadratic expressions with respect to the independent variables p_q and q . For other degrees of freedom the energy is expressed by one quadratic term with respect to the independent variable for each degree of freedom. The averaging of each quadratic term in the energy leads to the mean energy $\frac{1}{2}kT + \frac{1}{2}kT = kT$.

In the general case it can be said that each quadratic term entering into the energy of the system has a mean value equal to $\frac{1}{2}kT$. We have convinced ourselves of this by the example of monatomic and diatomic molecules. All our reasoning can be applied without any special difficulty to the case of polyatomic molecules.

Consider, for instance, triatomic molecules. A triatomic molecule can have a structure similar to that of the CO_2 molecule, or the H_2O and SO_2 molecules (fig. III.19). In the first case all the atoms are distributed along a line, and the molecule is called a linear molecule. Molecules of the second type are called non-linear molecules. In the case of a linear triatomic molecule, having nine degrees of freedom, the following forms of motion are possible: the translational motion of the molecule as a whole (three degrees of freedom), the rotation about two axes perpendicular to the axis of the molecule (two degrees of freedom), and the vibrational motion (four degrees of freedom). Possible types of vibrational motion for a linear molecule are shown in fig. III.19. The directions of motion in a given phase of normal vibrations are shown by arrows, while ν_1 , ν_2 and ν_3 denote vibrational frequencies. Two vibrations with frequency ν_2 taking place independently of each other in two perpendicular planes are possible. The mean energy of a linear molecule is made up of the mean energy of the translational, rotational and vibrational motions. Each of these forms of motion, as well as different normal vibrations, are independent of each other. Hence to each of these forms of motion taken separately we can apply the reasoning of the preceding paragraphs. To each degree of freedom of the translational and rotational motion there corresponds a mean energy $\frac{1}{2}kT$, while to each vibrational degree of freedom there corresponds an energy kT . Thus, the mean energy of a linear triatomic molecule is equal to

$$\bar{\epsilon} = 3 \frac{kT}{2} + 2 \frac{kT}{2} + 4kT = 6.5kT.$$

For a non-linear molecule the mean energy turns out to be different. For such a molecule the following forms of motion are possible: the translational motion of the molecule as a whole (three degrees of freedom), the rotation

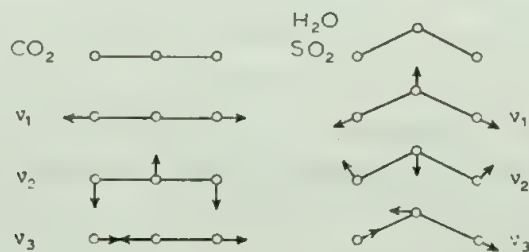


Fig. III.19

about three mutually perpendicular axes (three degrees of freedom), and the vibrations (three degrees of freedom). Possible normal vibrations are shown in fig. III.19. Everything else said about linear molecules holds also for non-linear triatomic molecules. Thus, the mean energy of a non-linear triatomic molecule is equal to

$$\bar{\epsilon} = 3 \frac{kT}{2} + 3 \frac{kT}{2} + 3kT = 6kT.$$

The problem of the mean energy of a polyatomic molecule can be considered in an analogous way. If the molecule contains n atoms, then out of $3n$ degrees of freedom there are always three translational degrees of freedom, three or two (in the case of a linear molecule) rotational degrees of freedom and respectively $3n - 6$ or $3n - 5$ vibrational degrees of freedom. Each of the degrees of freedom gives the corresponding contribution to the mean energy, the same as in the case of diatomic or triatomic molecules. Thus, the mean energy of a non-linear n -atomic molecule is equal to

$$\bar{\epsilon} = 3 \frac{kT}{2} + 3 \frac{kT}{2} + (3n-6) kT,$$

and that of a linear n -atomic molecule is

$$\bar{\epsilon} = 3 \frac{kT}{2} + 2 \frac{kT}{2} + (3n-5) kT.$$

In the general case one can write

$$\bar{\epsilon} = r \cdot \frac{1}{2} kT,$$

where r is the number of quadratic terms entering into the expression for the energy. Thus it turns out that all degrees of freedom of a molecule are equivalent: each quadratic term in the energy gives a contribution of $\frac{1}{2}kT$ to the mean energy of the molecule (the law of equipartition over degrees of freedom). The law of equipartition is a very general law. In deriving it we have not made any special assumptions, but have only assumed that the laws of statistical physics are valid and that the motion of the molecule obeys the laws of classical mechanics.

It should in addition be noted that the formulation of this important law is rather inappropriate, since it underlines the difference between vibrational and other degrees of freedom.

Since, however, the formulation presented is generally adopted, we shall make use of it in what follows. Moreover, we shall often for brevity call the number of quadratic terms in the energy the number of degrees of freedom.

Knowing the mean energy of a gas molecule, and taking into account that all molecules in an ideal gas are completely identical and equivalent, we can easily find the mean energy of the gas as a whole. If the gas has N molecules then the mean energy of the gas is equal to

$$E = N\bar{\epsilon} = N \cdot \frac{1}{2}rkT. \quad (39.26)$$

The heat capacity of a gas at constant volume C_V is equal to

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = \frac{1}{2}Nkr. \quad (39.27)$$

In particular, for one gram-mole of a gas the heat capacity at constant volume is equal to

$$C_V = \frac{1}{2}rR. \quad (39.28)$$

Correspondingly, the heat capacity at constant pressure is equal to

$$C_p = C_V + R = \frac{1}{2}(r+2)R. \quad (39.29)$$

Thus, the heat capacity of ideal gases turns out to be independent of the temperature, and is determined solely by the structure of the molecule. For monatomic gases the heat capacity at constant volume, calculated according to formula (39.29), is equal to $C_V = \frac{3}{2}R = 12.48 \text{ J mol}^{-1} \text{ K}^{-1}$.

For comparison with experiment, table 3 in §37 presents the measured values of the heat capacities of certain monatomic gases at constant volume and different temperatures. From table 3 it is seen that the theoretical predictions are well justified experimentally: the heat capacity of monatomic gases is constant over a wide temperature interval and has almost exactly the theoretical value.

For diatomic gases the situation is quite different. According to theoretical predictions, the heat capacity of diatomic gases should be equal to

$$C_V = \frac{7}{2}R \approx 29.11 \text{ J mol}^{-1} \text{ K}^{-1}. \quad (39.30)$$

However, experiment shows that diatomic gases do not in reality possess

such a large heat capacity. Moreover, it turns out that the heat capacity of diatomic gases depends on the temperature. This dependence is illustrated by fig. III.20. The general character of the dependence of the heat capacity on the temperature can be defined in the following way. Although at very high temperatures the heat capacity does not reach the theoretical value (39.30), it tends to it. As the temperature decreases the heat capacity also decreases and tends to the value

$$C_V = \frac{5}{2}R = 20.8 \text{ J mol}^{-1} \text{ K}^{-1} . \quad (39.31)$$

This value would be possessed by a diatomic molecule with an absolutely rigid link between the atoms, for which any vibrational motion is impossible. This vanishing of vibrational motion is completely inexplicable from the point of view of classical mechanics. From this standpoint, as we have stressed more than once, all degrees of freedom are completely equivalent. The vanishing of small vibrations as the temperature drops is in sharp contradiction with the basic propositions of classical mechanics. An even more striking example of such a contradiction is provided by the behaviour of hydrogen at

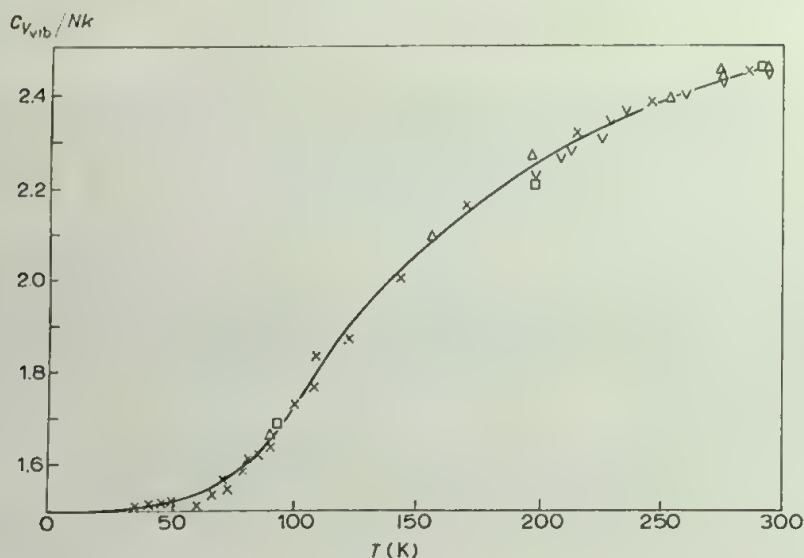


Fig. III.20

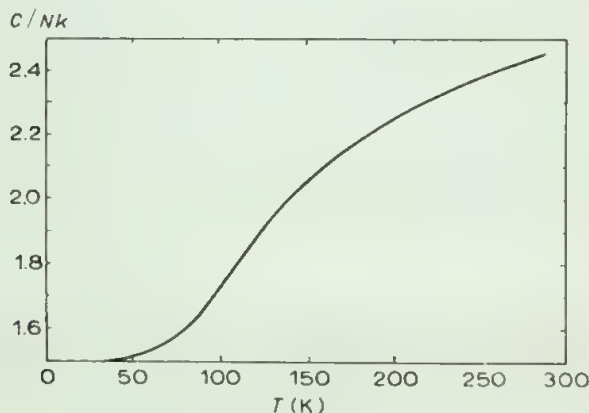


Fig. III.21

low temperatures. Namely, as is seen from fig. III.21, as the temperature decreases so does the heat capacity of hydrogen down to the value $\frac{3}{2}R$, equal to the value of the heat capacity of a monatomic gas. Thus, at low temperatures not only the vibrational motion but also the rotational motion vanishes in hydrogen molecules. A diatomic molecule can then perform only a translational motion.

From the point of view of our usual notions it is quite incomprehensible why an extended body, such as a diatomic molecule, can lose the ability to rotate. The contradiction between this fact and obvious ideas based on the laws of classical mechanics is even more evident than in the case of the vanishing of vibrations.

All that has been said about diatomic molecules holds also for polyatomic molecules. The part of the energy corresponding to vibrational degrees of freedom is always considerably smaller than that to be expected from the law of equipartition. For example, in the case of the linear molecule CO_2 the vibrational heat capacity should be equal to $4R \approx 32 \text{ J mol}^{-1} \text{ K}^{-1}$. In fact, it amounts to about $3.2 \text{ J mol}^{-1} \text{ K}^{-1}$ at room temperature, and increases up to a value of $24 \text{ J mol}^{-1} \text{ K}^{-1}$ at very high temperatures. Analogously, CH_4 molecules, possessing nine vibrational degrees of freedom to which, according to the law of equipartition, there should correspond a heat capacity of $72 \text{ J mol}^{-1} \text{ K}^{-1}$, have a vibrational heat capacity which does not exceed $13.8 \text{ J mol}^{-1} \text{ K}^{-1}$.

Thus, experiment points to the inapplicability of the law of equipartition

over degrees of freedom. But, as we have stressed, this law is based on only two assumptions: the assumption of the applicability of general statistical laws to simple molecular systems and the assumption of the applicability of the laws of classical mechanics to the description of the motion of individual molecules. Since the validity of the first assumption is beyond any doubt, the disagreement with experiment to which the law of equipartition leads shows that the second assumption is incorrect. In reality the motions of individual molecules obey the laws of quantum mechanics. The statistics of molecular systems moving according to the laws of quantum mechanics will be discussed below.

§40. The thermodynamic functions of a system which can be in two quantum states

Before going on to the consideration of more complex diatomic and polyatomic molecules, we have to consider the general aspect of the properties of a system which can be in two quantum states. We shall not specify the nature of these quantum levels. In the following paragraphs we shall see that these can be the quantum levels of the energy of rotational or vibrational motion. Sometimes they may also have a different nature.

Let us find the partition function of such a system. By definition,

$$z = \sum e^{-\epsilon_i/kT} g(\epsilon_i), \quad (40.1)$$

where ϵ_i are the quantum energy levels and $g(\epsilon_i)$ is the number of states of the particle whose energy is equal to ϵ_i . If g is different from unity, so that to one value of the energy of the system there correspond several different states, then the latter are called degenerate states, and their number is called the statistical weight of the energy level ϵ_i . In our case, where, for simplicity, we confine ourselves to two energy levels, the index i runs over the values 0, 1. We denote $g(\epsilon_0)$ by g_0 , and $g(\epsilon_1)$ by g_1 . Then

$$z = g_0 e^{-\epsilon_0/kT} + g_1 e^{-\epsilon_1/kT} = g_0 e^{-\epsilon_0/kT} \left(1 + \frac{g_1}{g_0} e^{-(\epsilon_1 - \epsilon_0)/kT} \right). \quad (40.2)$$

If the energy is expressed in thermal units of kT , then one can write that

$$\epsilon_1 - \epsilon_0 = kT_c, \quad (40.3)$$

where T_c is a certain temperature corresponding to the difference $\epsilon_1 - \epsilon_0$. By means of (40.3) the expression (40.2) can be written in the form

$$z = g_0 e^{-\epsilon_0/kT} \left(1 + \frac{g_1}{g_0} e^{-T_c/T} \right). \quad (40.4)$$

From the expressions (40.2) and (40.4) we see that, if the energy difference between the excited and ground state is so large that at a temperature T the inequality $\epsilon_1 - \epsilon_0 \gg kT$ or $T_c \gg T$ holds, then the second term in (40.4) can be disregarded, so that

$$z = g_0 e^{-\epsilon_0/kT}. \quad (40.5)$$

Physically this means that at a given temperature the probability for the system to get into an excited state with energy ϵ_1 is very small. The temperature is too low for thermal excitation to bring the system into an upper energy state with an appreciable probability. If, however, only one term enters into the partition function, so that the system is, with a probability equal to unity, in a state with an energy ϵ_0 , then its energy is exactly equal to ϵ_0 . This is confirmed by a direct calculation.

The partition function of a system consisting of N independent equal particles is

$$Z = \frac{1}{N!} (g_0 e^{-\epsilon_0/kT})^N.$$

According to (21.3), the energy of the system is equal to

$$E = kT^2 \frac{\partial \ln Z}{\partial T} = N\epsilon_0.$$

The heat capacity of the system at constant volume is

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = 0. \quad (40.6)$$

Thus, we see that at $T \ll T_c$ the presence of the second level does not affect the thermodynamic properties of the system at all, and it behaves as a system with constant energy. The heat capacity of the system at a sufficiently low temperature is equal to zero.

The behaviour, as the temperature rises, of the heat capacity of the system with two energy levels is of interest. If the inequality $T_c \gg T$ is not fulfilled, then one has to retain two terms in the partition function, writing

$$z = g_0 e^{-\epsilon_0/kT} \left(1 + \frac{g_1}{g_0} e^{-\Delta\epsilon/kT} \right).$$

In this case

$$Z = \frac{1}{N!} (g_0 e^{-\epsilon_0/kT})^N \left(1 + \frac{g_1}{g_0} e^{-\Delta\epsilon/kT} \right)^N.$$

The energy of the system is

$$E = kT^2 \frac{\partial \ln Z}{\partial T} = N\epsilon_0 + \frac{Ng_1 \Delta\epsilon e^{-\Delta\epsilon/kT}}{g_0 [1 + (g_1/g_0) e^{-\Delta\epsilon/kT}]}. \quad (40.7)$$

Finally, the heat capacity at a constant volume is

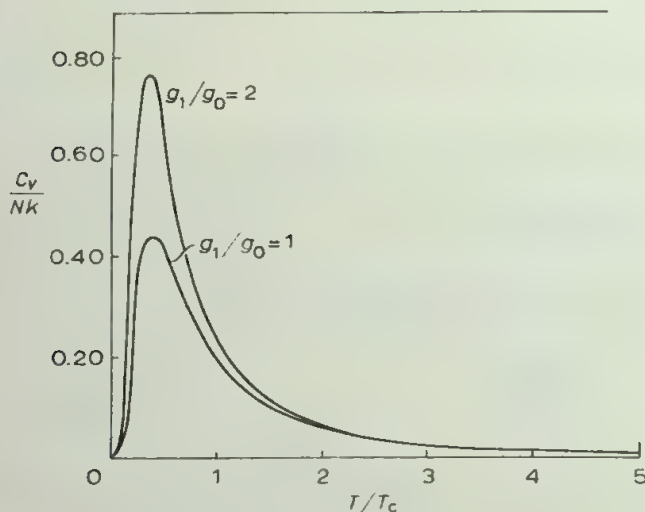


Fig. III.22

$$\begin{aligned}
 C_V &= \left(\frac{\partial E}{\partial T} \right)_V = N \left(\frac{g_1}{g_0} \right) \left(\frac{\Delta \epsilon}{kT} \right)^2 \frac{e^{-\Delta \epsilon/kT}}{[1 + (g_1/g_0) e^{-\Delta \epsilon/kT}]^2} = \\
 &= Nk \left(\frac{g_1}{g_0} \right) \left(\frac{T_c}{T} \right)^2 \frac{e^{-T_c/T}}{[1 + (g_1/g_0) e^{-T_c/T}]^2}. \quad (40.8)
 \end{aligned}$$

The form of the heat capacity is shown in fig. III.22. From fig. III.22 it is seen that this form is unusual: at $T = 0$ the heat capacity is zero, in agreement with what was said earlier. At higher temperatures the heat capacity increases and has a characteristic maximum. As the temperature rises further the heat capacity again reduces to zero. This latter fact represents a characteristic feature of a system with two levels. The reason why C_V reduces to zero is understandable from formula (40.7). At a very high temperature the energy of the system is equal to

$$E \approx N\epsilon_0 + \frac{Ng_1}{g_0} \frac{\epsilon_1 - \epsilon_0}{1 + g_1/g_0} = \text{const} \quad (40.9)$$

and does not depend on temperature. Physically this means that at $T \gg T_c$ the thermal excitation is so large that the system may with equal ease be in the ground state or in an excited state. The probability for the system to be in an excited state is comparable to the probability of being in the ground state.

If the system had other excited levels, ranging up to very large energies, this would not hold. Even at a high temperature there would be energy levels which the system would only have a small probability of occupying. Hence the mean energy of such a system would not be expressed by a formula of the type of (40.9) and, at a high temperature, would vary with temperature.

Correspondingly, the heat capacity of the system would not reduce to zero at a high temperature.

The form of the heat capacity, with a maximum and the reduction to zero on the low and high temperatures sides, is typical of a system with levels lying in a finite energy interval. The presence of two levels simplifies the calculation, but it is not essential. A similar form of the heat capacity also occurs in systems with several levels. It is of importance only that they lie sufficiently close to each other, so that the temperature at which the condition $kT \gg \epsilon_0 - \epsilon_n$ is fulfilled may be reached. A typical example of atoms with two closely lying levels are atoms of the halogens and the alkali metals. In the case of halogens the lowest level is four-fold degenerate, $g_0 = 4$, and the excited level lying closest to it is two-fold degenerate, $g_1 = 2$. The

distance between the levels is $T_c = 582.7$ K for fluorine, $T_c = 1299$ K for chlorine, and $T_c = 5275$ K for bromine. The next energy level lies much higher: the corresponding temperature amounts to several tens of thousands of degrees (for example, for bromine it amounts to about 88×10^3 K) and gives practically no contribution to the heat capacity.

§41. Diatomic molecules

The simplest molecules are diatomic molecules, representing a stable combination of two identical or dissimilar atoms. We cannot here discuss in detail the problem of the nature of the forces leading to the formation of molecules from free atoms nor describe in detail the motion of atoms in molecules. We shall confine ourselves to the most superficial characteristics of molecules, presenting only those data which we shall need in what follows. (Ch. 10 of Part V is devoted to the theory of molecules.)

Fig. III.23 shows a typical curve representing the energy of interaction of the electron shells of atoms as a function of the distance between them. The potential energy of interaction has a minimum at a certain point denoted by r_0 . To the right of it at large distances the tangent to the curve and, consequently, the force of interaction also is positive. This means that the atoms

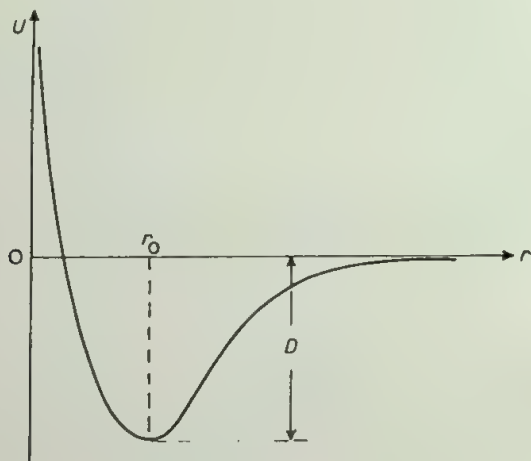


Fig. III.23

attract each other. To the left of the point r_0 , in the region where the electron shells are overlapping, a strong repulsion between the atoms arises.

Thus there corresponds to the position of stable equilibrium in a molecule a definite distance between the nuclei of the atoms, which can be called the diameter of the molecule (see table 4 below). The energy $U(r_0)$ of the electron shells has a minimum value. If the distance between the nuclei changes by a small value x , then the energy of the molecule becomes equal to $U(r_0+x)$. For small values of x it can be expanded in a series in powers of x and one can restrict oneself to the first terms of the expansion:

$$\begin{aligned} U(r_0+x) &\approx U(r_0) + \left(\frac{\partial U}{\partial r}\right)_{r=r_0} x + \left(\frac{\partial^2 U}{\partial r^2}\right)_{r=r_0} \frac{1}{2}x^2 = \\ &= U(r_0) + \frac{1}{2}\kappa x^2 \end{aligned} \quad (41.1)$$

(at the point of the minimum the first derivative is equal to zero, and the second derivative is positive). Formula (41.1) shows that when the atoms depart from the equilibrium position they are acted upon by a quasi-elastic force bringing them back into the equilibrium position. From this it is clear that in a molecule, in addition to the motion of electrons in atomic shells, the vibration of the atoms about the equilibrium position is also possible. Moreover, a molecule as a whole can rotate about two axes perpendicular to the straight line connecting the nuclei.

Thus, the energy of a molecule can be considered as made up of the energy of translational motion of the molecule as a whole in space, the energy of motion of the electrons, the vibrational energy and the energy of rotation of the molecule. The translational motion of a diatomic molecule differs in no way from that of a monatomic molecule. Hence we shall be interested only in the internal motion of a diatomic molecule. Its internal energy can be written in the form

$$\epsilon_{\text{int}} = \epsilon_{\text{el}} + \epsilon_{\text{vib}} + \epsilon_{\text{rot}}, \quad (41.2)$$

where ϵ_{el} is the energy of motion of the electrons, ϵ_{vib} is the energy of vibrations, and ϵ_{rot} is the energy of rotation. The internal motion of the molecule turns out to be quantized. The energies ϵ_{el} , ϵ_{vib} and ϵ_{rot} take on a discrete series of values. The spacing $\Delta\epsilon_{\text{el}}$ between neighbouring energy levels of the electrons in the molecule turns out to be much larger than the spacing $\Delta\epsilon_{\text{vib}}$ between neighbouring energy levels of the vibrational motion. In its turn the spacing $\Delta\epsilon_{\text{vib}}$ between neighbouring energy levels of the vibrational motion is very large in comparison with the spacing $\Delta\epsilon_{\text{rot}}$ between neighbouring energy levels of the rotational motion. Thus,

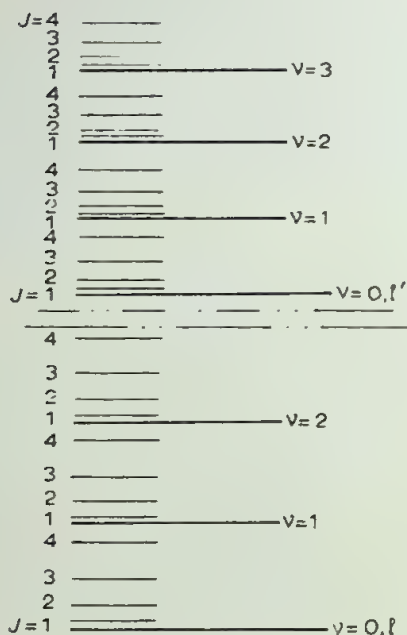


Fig. III.24

$$\Delta\epsilon_{el} \gg \Delta\epsilon_{vib} \gg \Delta\epsilon_{rot} \quad (41.3)$$

Hence the energy levels of a molecule are distributed in the way shown schematically in fig. III.24. In this drawing the vibrational and rotational levels pertaining to two electron levels, l and l' , are presented schematically. The spacing between the latter is large and is represented by dashed-dotted lines to show that there is not sufficient room for it in the scale of the drawing. Bold lines correspond to vibrational states with quantum numbers $\nu = 0, 1, 2$. Fine lines correspond to different rotational levels with quantum numbers $J = 1, 2, 3, 4$. Since the motion of electrons is much faster than that of heavy nuclei (for the vibrations and rotation of the molecule as a whole), it can be assumed in the first approximation that the motion of the nuclei does not affect that of the electrons.

Further, if the amplitude of vibration of the nuclei of the molecule is sufficiently small, the effect of the vibrational motion on the rotation can be disregarded. For a small vibrational amplitude the change in the distances between the nuclei is so small that the corresponding change in the moment

of inertia of the molecule is very small and can be ignored. The rotation will proceed with a constant moment of inertia, as though there were no vibrations. Thus, to a first approximation all three forms of motion in a molecule can be considered to be independent of each other.

It should be noted that the accuracy of contemporary methods of measurement is such that for many purposes the calculations based on the concept of the independent rotational and vibrational motion turn out to be insufficiently accurate. In contemporary theory one has to take into account the change in the moment of inertia of the molecule due to its vibration.

The spacing between neighbouring energy levels of the electron motion, as between energy levels in atoms, is of the order of several electron-volts, which corresponds to a temperature of several thousand degrees. In order to bring a molecule from one electronic level to another the corresponding energy must be imparted to it. This is only possible at very high temperatures (or for non-thermal effects on the molecule, for example, when it is irradiated by light, bombarded by fast electrons, etc.). As a rule, however, it can be assumed that these sources of excitation of the electronic motion are absent and that molecules are in the lowest electron energy level. In what follows we shall confine ourselves to the study of this case. Thus, in considering the thermal motion of molecules it is possible to disregard electron energy levels.

We shall consider now the vibrational motion of a diatomic molecule. The oscillations of the two nuclei about the equilibrium distance can be reduced to the oscillatory motion of one material point with a reduced mass $\mu = m_1 m_2 / (m_1 + m_2)$. Such a material point represents a linear oscillator, considered in §1. For a sufficiently small vibration amplitude it can be considered to be a harmonic oscillator. The energy of a harmonic oscillator takes on a discrete series of values given by formula (1.18):

$$\epsilon_{\text{vib}} = h\nu(n + \frac{1}{2}),$$

where the quantum number n takes on a series of integer values: $n = 0, 1, 2, \dots$; ν is the classical frequency connected with the constant of the quasi-elastic force κ and the mass of the oscillator by the usual relation

$$\nu = (\kappa\mu^{-1})^{\frac{1}{2}}/2\pi.$$

All levels of the oscillator are non-degenerate, so that to each value of the quantum number n there corresponds a definite energy ϵ_{vib} . The energy difference between neighbouring levels of the vibrational motion is equal to

$$\Delta\epsilon_{\text{vib}} = h\nu(n+\frac{1}{2}) - h\nu(n-1+\frac{1}{2}) = h\nu$$

and does not depend on the quantum number n ; the energy levels are distributed at equal distances from each other.

According to the Bohr frequency rule, when a system passes over from one energy level to another a photon with energy $h\nu$ is emitted or absorbed. In quantum mechanics it turns out that the change in the quantum number n obeys the so-called selection rule:

$$\Delta n = \pm 1.$$

Measuring the frequencies of the light absorbed or emitted by the molecules, one can determine the natural frequency ν of the molecule and find the constant of the quasi-elastic force κ and the energy difference $\Delta\epsilon_{\text{vib}}$. The values of these quantities for several molecules are presented in table 4. The radiation emitted or absorbed by molecules as their vibrational state is changed (for a fixed value of the electron energy) lies in the infrared spectral region *, as a rule between 100 and 4000 cm^{-1} .

Table 4
Basic quantities characterizing the properties of diatomic molecules

Molecule	Distance between the atoms r_0 (10^{-8} cm)	Moment of inertia I (10^{-40} g cm^2)	Vibrational frequency ν/c (cm^{-1})	Rotational constant $\frac{B}{c} = \frac{h}{8\pi^2 Ic}$ (cm^{-1})	Constant of the quasi- elastic force (10^{-5} dyne cm)	Dissociation energy D (eV)
H ₂	0.74	0.46	4276	59.35	5.1	4.48
N ₂	1.10	13.84	2360	2.00	22.2	7.38
O ₂	1.21	19.13	1580	1.45	11.3	5.08
Cl ₂	1.99	113.5	565	0.24	3.21	2.47
HCl	1.27	2.67	2989	10.6	8.65	4.40
CO	1.13	14.37	2169	1.92	18.6	9.61
NO	1.15	16.43	1906	1.68	15.4	5.29

The formulae given for the energy of a vibrating molecule are valid only in the approximation of small vibrations. For a high excitation (for example,

* It should be noted that this does not refer to symmetric molecules of the type of H₂ and O₂, for which there are no such transitions. For these molecules $\Delta\epsilon_{\text{vib}}$ is determined from transitions with a simultaneous change of electron states.

at a high temperature) their amplitude becomes large, and one must take into account anharmonic terms in the potential energy.

We shall now consider the rotational motion of a diatomic molecule. If the change in the moment of inertia of the molecule due to vibrations is disregarded, then the molecule can be considered as a rigid rotator with a moment of inertia $I = m_1 m_2 r_0^2 / (m_1 + m_2)$ rotating about the centre of mass. As is shown in quantum mechanics (Part V, §81), the energy of a rotating body is expressed by the formula

$$\epsilon_{\text{rot}} = \frac{h^2}{8\pi^2 I} J(J+1) = hBJ(J+1), \quad (41.4)$$

where J is a quantum number taking on integer values: $J = 0, 1, 2, \dots$, and $B = h^2/8\pi^2 I$ is a constant called the rotational constant. It turns out that the states of a rotator corresponding to given rotational energy are $(2J+1)$ -fold degenerate (see §30 and §81 of Part V).

When the quantum states of the molecule change, the quantum number changes by an amount $\Delta J = \pm 1$. The spacing between neighbouring rotational energy levels is equal to

$$\Delta\epsilon_{\text{rot}} = \frac{2h^2}{8\pi^2 I} (J+1).$$

By observing the radiation in the transition between rotational levels, one can determine the value of $\Delta\epsilon_{\text{rot}}$ and, consequently, the moment of inertia I of the molecule. The value of these quantities for some molecules can be calculated by means of table 4. Substituting into the expressions for ϵ_{rot} and ϵ_{vib} the values of the constants given in the fourth and fifth columns of the table, we verify the validity of the assumption that $\Delta\epsilon_{\text{vib}} \gg \Delta\epsilon_{\text{rot}}$ ($\Delta\epsilon_{\text{rot}}$ is smaller than $\Delta\epsilon_{\text{vib}}$ by a factor of 800–1000). It should be noted that in practice one seldom manages to observe the transitions of a molecule between different rotational levels for an unchanged electron state and an unchanged vibrational state, since $\Delta\epsilon_{\text{rot}}$ is so small that the corresponding frequencies $\nu = \Delta\epsilon_{\text{rot}}/h$ lie in the far infrared spectral region, where the accuracy of measurements is poor. The emission spectra and absorption spectra of molecules in the visible spectra region are those most often observed. These spectra arise when the vibrational, rotational and electronic state of a molecule are changed simultaneously. The emission (or absorption) spectrum has in this case the character of a group of close spectral lines merging in a low-resolution spectroscopy into continuous bands (molecular

band spectrum). The origin of the bands is easily understood from fig. III.24. Let, for example, a transition take place from upper levels to the lowest level. The fundamental frequency is emitted in the transition from the level 2, $\nu = 0$, $J = 1$. Frequencies close to the fundamental frequency are emitted in the transition from the levels 2, $\nu = 1$, $J = 0$; $\nu = 2$, $J = 0$ and so on. Thus, when the vibrational, rotational and electronic states of a molecule are changed simultaneously a number of frequencies lying close to each other are emitted (since the inequality (41.3) is fulfilled). The totality of spectroscopic data has made it possible to establish the position of energy levels for a very large number of diatomic molecules.

§42. Thermodynamic functions of diatomic gases

We can now pass on to the consideration of the heat capacities and thermodynamic functions of diatomic gases, the calculation of which was insuperably difficult for classical statistics. The method of calculation of the thermodynamic functions for diatomic gases in no way differs from that which we have already considered for monatomic gases. Since the molecules are identical, independent particles, the partition function for an entire gas containing N molecules can be written in the form

$$Z = \frac{1}{N!} (z)^N. \quad (42.1)$$

We have to find the partition function for one molecule. The energy of a molecule can be divided into the energy of its motion as a whole in space and the energy of its internal motion:

$$\epsilon = \epsilon_{\text{transl}} + \epsilon_{\text{int}}.$$

Since these two forms of motion are independent, the number of states of the system corresponding to an energy ϵ is resolved into the number of states corresponding to the energy of translational motion and the number corresponding to the internal motion:

$$\Omega = \Omega(\epsilon_{\text{transl}}) \Omega(\epsilon_{\text{int}}).$$

Accordingly, the partition function can be resolved into the product of two factors:

$$\begin{aligned}
 z &= \sum \exp(-\epsilon_{\text{transl}}/kT) \Omega(\epsilon_{\text{transl}}) \cdot \sum \exp(-\epsilon_{\text{int}}/kT) \Omega(\epsilon_{\text{int}}) = \\
 &= z_{\text{transl}} z_{\text{int}},
 \end{aligned}
 \quad (42.2)$$

where z_{transl} is the partition function associated with the translational motion of the molecule as a whole, and z_{int} is the partition function for the internal motion.

The partition function for the translational motion of a diatomic molecule is no different from that of a monatomic molecule, since it moves in space as a material point with a mass $m = m_1 + m_2$ located at the centre of mass of the molecule. Hence for the partition function of translational motion one can write that

$$z_{\text{transl}} = \left(\frac{2\pi mkT}{h^2} \right)^{\frac{3}{2}} V. \quad (42.3)$$

The calculation of z_{int} is more complicated. The internal motion of a diatomic molecule reduces to its rotation with respect to two mutually perpendicular axes and the vibrations of the atoms about the equilibrium position. In the first approximation one can neglect the effect of small vibrations on the value of the moment of inertia of the molecule and assume the vibrational and rotational motions to be independent of each other (see §41). One can disregard the electron energy since it remains unchanged. Hence, according to (41.2), the energy of internal motion of the molecule can be written in the form

$$\epsilon_{\text{int}} = \epsilon_{\text{vib}} + \epsilon_{\text{rot}}.$$

Accordingly the partition function for the internal motion resolves into the product of two factors:

$$\begin{aligned}
 z_{\text{int}} &= \sum \exp [-(\epsilon_{\text{vib}} + \epsilon_{\text{rot}})/kT] \Omega(\epsilon_{\text{vib}}) \Omega(\epsilon_{\text{rot}}) = \\
 &= \sum \exp(-\epsilon_{\text{vib}}/kT) \Omega(\epsilon_{\text{vib}}) \cdot \sum \exp(-\epsilon_{\text{rot}}/kT) \Omega(\epsilon_{\text{rot}}) = \\
 &= z_{\text{vib}} z_{\text{rot}}.
 \end{aligned}
 \quad (42.4)$$

Substituting the expressions for z from (42.2) and (42.4) into (42.1), we obtain

$$Z = \frac{1}{N!} (z_{\text{transl}})^N (z_{\text{vib}})^N (z_{\text{rot}})^N. \quad (42.5)$$

From this the expressions for the thermodynamic functions can be found:

$$\begin{aligned} F &= -kT \ln Z = \\ &= -kT \ln \frac{z_{\text{transl}}^N}{N!} - kT \ln z_{\text{rot}}^N - kT \ln z_{\text{vib}}^N = \\ &= F_{\text{transl}} + F_{\text{vib}} + F_{\text{rot}}, \end{aligned} \quad (42.6)$$

where F_{transl} , F_{vib} and F_{rot} denote the individual components of the free energy due to the translational, vibrational and rotational motions of the gas molecules. The expression for F_{transl} is the same as the free energy of a monatomic gas [formula (37.7)] if in the latter the mass of one atom is replaced by the total mass of the diatomic molecule. Analogously,

$$E = kT^2 \frac{\partial \ln Z}{\partial T} = E_{\text{transl}} + E_{\text{vib}} + E_{\text{rot}}, \quad (42.7)$$

$$S = - \left(\frac{\partial F}{\partial T} \right)_V = S_{\text{transl}} + S_{\text{vib}} + S_{\text{rot}}, \quad (42.8)$$

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = C_{V_{\text{transl}}} + C_{V_{\text{vib}}} + C_{V_{\text{rot}}}. \quad (42.9)$$

Thus, all the thermodynamic functions are resolved into individual components. Each component corresponds to one of the independent forms of motions of the diatomic molecule: the translational, vibrational or rotational motion. In order to calculate the thermodynamic functions of a diatomic gas it is necessary to find the corresponding partition functions of the internal motion. In the following sections we shall consider the partition functions for the vibrational and rotational motions.

§43. The vibrational partition function and the contribution of vibrations to the energy and heat capacity

In the first approximation a vibrating diatomic molecule can be considered as a quantum harmonic oscillator the energy of which is expressed by formula

(1.18). All energy levels of the oscillator are non-degenerate, i.e. with a weight $\Omega = 1$. Substituting the expression for the energy (1.18) into the partition function, we have

$$z_{\text{vib}} = \sum_{n=0}^{\infty} \exp \left[-h\nu \left(n + \frac{1}{2} \right) / kT \right] = e^{-h\nu/2kT} \sum_{n=0}^{\infty} e^{-h\nu n/kT}. \quad (43.1)$$

Making use of the well-known formula for the sum of an infinitely decreasing geometric series, we obtain

$$z_{\text{vib}} = e^{-h\nu/2kT} \sum_{n=0}^{\infty} (e^{-h\nu/kT})^n = \frac{e^{-h\nu/2kT}}{1 - e^{-h\nu/kT}}. \quad (43.2)$$

From formula (43.2) we see that the partition function and, consequently, the thermodynamic quantities also are determined by the value of the variable $h\nu/kT$. In the notation of §40 we can, expressing $h\nu$ in energy units, write $h\nu = kT_c$, where T_c is the so-called characteristic temperature. Formula (43.2) can be rewritten in the form

$$z_{\text{vib}} = \frac{e^{-T_c/2T}}{1 - e^{-T_c/T}}.$$

Now we calculate the thermodynamic functions of a diatomic molecule corresponding to its vibrational motion. We first of all find the mean vibrational energy:

$$\begin{aligned} E_{\text{vib}} &= N\bar{\epsilon}_{\text{vib}} = NkT^2 \frac{\partial}{\partial T} \ln z_{\text{vib}} = \\ &= NkT^2 \frac{\partial}{\partial T} \ln \frac{e^{-T_c/2T}}{1 - e^{-T_c/T}} = \\ &= \frac{NkT_c}{2} \coth \frac{T_c}{2T} = \frac{Nh\nu}{2} \coth \left(\frac{h\nu}{2kT} \right). \end{aligned} \quad (43.3)$$

For the vibrational heat capacity we find

$$\begin{aligned}
 C_{V_{\text{vib}}} &= \frac{\partial E_{\text{vib}}}{\partial T} = \\
 &= \frac{Nk}{4} \left(\frac{h\nu}{kT} \right)^2 \frac{1}{\sinh^2(h\nu/2kT)} = \\
 &= \frac{Nk}{4} \left(\frac{T_c}{T} \right)^2 \frac{1}{\sinh^2(T_c/2T)}. \quad (43.4)
 \end{aligned}$$

We see that the mean vibrational energy and heat capacity turn out to be complicated functions of the temperature T and the characteristic temperature T_c (or the natural frequency ν). Consider the limiting form of these functions at high temperatures ($T \gg T_c$) and low temperatures ($T \ll T_c$). In the first case the exponential function can be expanded in a series, and one can restrict oneself to the first terms of the expansion. This gives

$$E_{\text{vib}} \approx NkT, \quad (43.5)$$

$$C_{V_{\text{vib}}} \approx Nk. \quad (43.6)$$

At low temperatures $e^{T_c/T} \gg 1$, so that

$$E_{\text{vib}} \approx \frac{1}{2}NkT_c + NkT_c e^{-T_c/T}, \quad (43.7)$$

$$C_{V_{\text{vib}}} \approx Nk (T_c/T)^2 e^{-T_c/T}. \quad (43.8)$$

Formulae (43.5) and (43.6) agree with the classical formulae of §39. On the contrary, at low temperatures the expressions for the energy and heat capacity differ fundamentally from the classical ones. As the temperature decreases the vibrational energy tends to a constant limit $E_0 = \frac{1}{2}NkT_c = \frac{1}{2}Nh\nu$. The latter quantity, representing the vibrational energy of the molecules at absolute zero, is called the zero point energy. The existence of the zero point energy is a characteristic feature of quantum motion. It is an expression of the fact that in quantum theory the concept of a particle at rest is devoid of any physical meaning.

The numerical value of E_0 can be found from spectroscopic data. For this it is only necessary to find the value of the natural vibrational frequency ν of the molecule. The heat capacity at low temperatures turns out to be a small quantity decreasing according to an exponential law, i.e. tending to zero as $T \rightarrow 0$. Thus, the general scheme of the variation of the energy and heat

capacity with the temperature reduces to the fact that at high temperatures, when the thermal energy kT is large in comparison with the spacing $\Delta\epsilon = h\nu = kT_c$ between the energy levels, the heat capacity and the energy are given by the classical expressions; at low temperatures the energy tends to a limiting value — to the zero point energy of the quantum oscillator, while the heat capacity tends to zero. Such behaviour of these quantities is in agreement with general considerations. At high temperatures the magnitude of the quantum steps $\Delta\epsilon$ is small in comparison with the thermal energy, so that the oscillator can be in a large number of excited quantum states, and its energy can be assumed to vary continuously, as in the case of a classical oscillator. On the contrary, at low temperatures the oscillator is always in the ground state and thermal excitation is insufficient to bring it into upper excited states.

The curve showing the dependence of the mean energy of the oscillator on the ratio T/T_c , given by formula (43.3), is shown in fig. III.25. From fig. III.25 it is seen that, when T approaches T_c , a smooth transition takes place between the limiting values (43.5) and (43.7). The basic difference between the classical expression for the mean energy of the oscillator (43.5) and the quantum expression (43.3) lies in the fact that in the latter case the energy depends on the frequency. Because of this the temperature still does not completely characterize the energy of the oscillator. For the same temperature two oscillators with different natural oscillation frequencies will have different energies. Fig. III.26 shows the dependence of the energy of the oscillator on the frequency for a fixed temperature T .

The heat capacity decreases smoothly with decreasing temperature from its classical value (43.6) down to zero. Thus, the vanishing of the vibrational

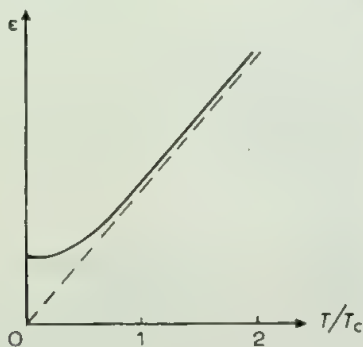


Fig. III.25

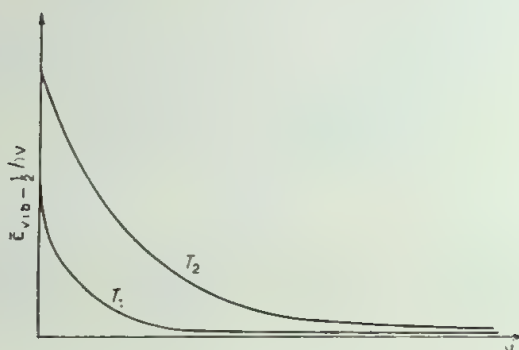


Fig. III.26

heat capacity (the “freezing” of oscillations mentioned in §39) appears in the most direct and natural way in considering the properties of a molecule as a quantum oscillator.

At low temperatures the vibrational frequency ν turns out to be relatively very large in comparison with kT/h . There corresponds to a high frequency a great rigidity of the bond between the two atoms. With decreasing temperature the increase of the relative rigidity leads to a progressive damping of oscillations.

So that one may have an idea of the order of magnitude of the quantities and, in particular, that of the characteristic temperatures of different molecules, we give in table 5 the corresponding values for a number of molecules. The values of the natural vibrational frequencies of the molecules are found from spectroscopic data.

Table 5

Molecule	Characteristic temperature (10^3 K)	Molecule	Characteristic temperature (10^3 K)
H ₂	6.0	HCl	4.14
N ₂	3.34	HBr	3.7
O ₂	2.23	HI	3.2
CO	3.07		

From table 5 the following conclusion, important from the practical point of view, can be drawn directly: since the characteristic temperatures of vibrational motion for all the molecules are of the order of several thousand degrees, a temperature of the order of 300 K corresponds to the limiting case $T \ll T_c$. Hence the vibrational heat capacity is very small for most molecules at room temperature. For example, in the case of HI at 640 K a calculation according to the general formula (43.4) shows that the vibrational heat capacity amounts to about $0.33 \text{ J mol}^{-1} \text{ K}^{-1}$. In most cases of practical importance at not very high temperatures it can be assumed that the vibrational motion is damped and that its contribution to the heat capacity is zero. In any case it is substantially smaller than that to be expected from the law of equipartition. The vibrational part of the heat capacity depends on the temperature and is different for molecules of different substances.

We now consider the calculation of other thermodynamic quantities. The free energy due to the vibrational motion has the form

$$\begin{aligned} F_{\text{vib}} &= -NkT \ln z_{\text{vib}} = \frac{1}{2}Nh\nu + NkT \ln(1 - e^{-h\nu/kT}) = \\ &= \frac{1}{2}NkT_c + NkT \ln(1 - e^{-T_c/T}). \end{aligned} \quad (43.9)$$

Correspondingly, the entropy is

$$\begin{aligned} S_{\text{vib}} &= -\frac{\partial F_{\text{vib}}}{\partial T} = -Nk \ln(1 - e^{-h\nu/kT}) + \frac{Nh\nu}{T} \frac{1}{e^{h\nu/kT} - 1} = \\ &= Nk \left(\frac{T_c}{T} \right) \frac{1}{e^{T_c/T} - 1} - Nk \ln(1 - e^{-T_c/T}). \end{aligned} \quad (43.10)$$

At high temperatures one can carry out the expansion in powers of the ratio T_c/T and confine oneself to the first terms of the expansion. This gives

$$F_{\text{vib}} \approx E_0 + NkT \ln(T_c/T),$$

$$S \approx Nk - Nk \ln(T_c/T).$$

On the contrary, at low temperatures $e^{-T_c/T} \ll 1$, so that

$$F_{\text{vib}} \approx E_0, \quad S_{\text{vib}} \approx 0.$$

For the practical calculation of the partition function of the vibrational

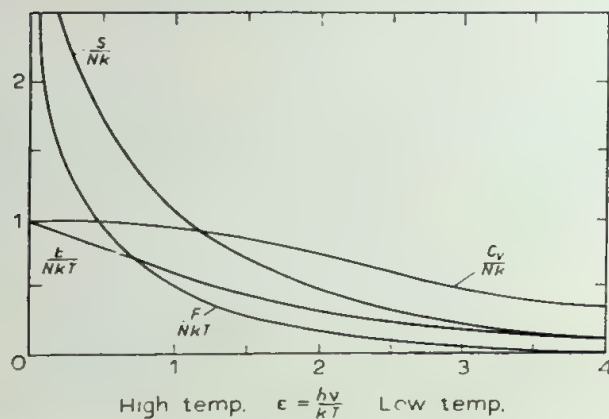


Fig. III.27

motion and the thermodynamic quantities it is necessary to know a characteristic molecular constant: the natural vibrational frequency ν of the molecule. Its value for most diatomic molecules is known from spectroscopic data, in particular from infrared vibrational spectra.

The values of the functions involved in formulae (43.3), (43.4), (43.9) and (43.10) are tabulated and are found directly from tables. Their dependence on the ratio $h\nu/kT$ is shown in fig. III.27.

§44. The rotational partition function and the contribution of rotation to thermodynamic functions

We shall consider now the partition function for the rotational motion of a diatomic molecule.

The energy of the rotator takes on a discrete series of values

$$\epsilon_{\text{rot}} = \frac{h^2}{8\pi^2 I} J(J+1), \quad (44.1)$$

where $J = 0, 1, 2, 3, \dots$

Each state with a definite rotation energy, i.e. with a definite value of the rotation quantum number, turns out to be $(2J+1)$ -fold degenerate. Hence the partition function of the rotational motion has the form

$$z_{\text{rot}} = \sum_{J=0}^{\infty} (2J+1) \exp\left(-\frac{h^2 J(J+1)}{8\pi^2 I k T}\right) \quad (44.2)$$

The partition function z_{rot} depends on the ratio $h^2/8\pi^2 I k T = T_c/T$, where $T_c = h^2/8\pi^2 I k$ is the characteristic temperature for the rotation. Thus,

$$z_{\text{rot}} = \sum_{J=0}^{\infty} (2J+1) \exp\left(-\frac{T_c J(J+1)}{T}\right). \quad (44.3)$$

Table 6 gives the values of the characteristic temperatures for the rotation of different diatomic molecules.

Table 6

Molecule	Characteristic temperature (K)	Molecule	Characteristic temperature (K)
H ₂	85.4	O ₂	2.07
D ₂	43	HCl	15.1
N ₂	2.85	HI	9.0

From table 6 it is seen that, in contrast to the characteristic temperatures for the vibrational motion, the characteristic temperatures for the rotational motion are extremely small and lie considerably below the condensation point of a gas at normal pressure. Exceptions to the rule are the molecules H₂ and D₂, for which the characteristic temperatures are relatively large and lie above the condensation temperatures. The high characteristic temperatures of H₂ and D₂ are due to the smallness of their moments of inertia ($I(\text{H}_2) = \frac{1}{2}m_{\text{H}}a^2$ and $I(\text{D}_2) = \frac{1}{2}m_{\text{D}}a^2$, where m_{H} and m_{D} are the masses of the proton and deuteron, and a is the distance between them in the molecules). Hence for all molecules, except H₂ and D₂, it can be assumed that the spacing between two successive rotational energy levels is small in comparison with the thermal energy. In other words, the temperature is always high with respect to the rotation of heavy molecules.

For high temperatures the summation over individual energy levels in (44.3) can be replaced by integration over the almost continuous levels:

$$\begin{aligned}
 z_{\text{rot}} &= \sum (2J+1) \exp\left(-\frac{T_c J(J+1)}{T}\right) \approx \\
 &\approx \int_0^{\infty} (2J+1) \exp\left(-\frac{T_c J(J+1)}{T}\right) dJ.
 \end{aligned}
 \quad (44.4)$$

Introducing a new integration variable, $y = J(J+1)$, we find

$$z_{\text{rot}} = \int_0^{\infty} \exp\left(-\frac{T_c y}{T}\right) dy = \frac{T}{T_c} = \frac{8\pi^2 I k T}{h^2}. \quad (44.5)$$

We thus arrive at the classical expression for z_{rot} .

In the next approximation the partition function can be calculated by means of the well-known Euler summation formula*:

$$\sum_{J=0}^{\infty} f(J) \approx \int_0^{\infty} f(x) dx + \frac{1}{2}f(0) - \frac{1}{12}f'(0) + \frac{1}{720}f'''(0).$$

In the given case $f(J) = (2J+1) \exp[-T_c J(J+1)/T]$, so that

$$f(0) = 1; \quad f'(0) = 2 - (T_c/T); \quad f'''(0) \approx -12 T_c/T$$

and for z_{rot} we obtain

$$z_{\text{rot}} \approx \frac{8\pi^2 I k T}{h^2} + \left(\frac{1}{3} + \frac{h^2}{15 \times 8\pi^2 I k T} \right). \quad (44.6)$$

The second and third terms represent the quantum correction to the classical value of z_{rot} . From formula (44.6) it is seen that this correction is small at a temperature T higher than T_c , and that it decreases rapidly with increasing T .

Taking into account the quoted values of the characteristic temperatures, it can be said that the entire range of realistic temperatures lies much higher

* See, for example, A. Gelfond, *Ischislenie konechnykh raznostei (Calculation of finite differences)*, (Gostekhizdat, Moscow, 1952) p. 343. H. Margenau and G. M. Murphy, *Mathematics of physics and chemistry* (D. Van Nostrand, New York, 1948).

than T_c . Hence the quantum corrections to z_{rot} for all except the lightest molecules, play a negligible role.

For low temperatures ($T \ll T_c$) in the general expression for z_{rot} only the first, that is the largest terms should be retained. This gives

$$z_{\text{rot}} \approx 1 + 3 \exp(-2T_c/T). \quad (44.7)$$

Now we find the terms of the thermodynamic functions associated with the rotational motion. Obviously, we have

$$\begin{aligned} E_{\text{rot}} &= NkT^2 \frac{\partial \ln z_{\text{rot}}}{\partial T} = \\ &= NkT^2 \frac{\partial}{\partial T} \ln \sum_{J=0}^{\infty} (2J+1) \exp\left(-\frac{T_c J(J+1)}{T}\right). \end{aligned} \quad (44.8)$$

For high temperatures we have

$$E_{\text{rot}} \approx NkT \left(1 - \frac{h^2}{24\pi^2 I k T}\right). \quad (44.9)$$

Correspondingly, the rotational heat capacity at high temperatures has the classical value:

$$C_{V_{\text{rot}}} \approx Nk. \quad (44.10)$$

At low temperatures

$$\begin{aligned} E_{\text{rot}} &= NkT^2 \frac{\partial}{\partial T} \ln \left[1 + 3 \exp\left(-\frac{h^2}{4\pi^2 I k T}\right) \right] \approx \\ &\approx \frac{3h^2 N}{4\pi^2 I} \exp\left(-\frac{h^2}{4\pi^2 I k T}\right) \end{aligned} \quad (44.11)$$

and the heat capacity is

$$C_{V_{\text{rot}}} \approx 3 \left(\frac{h^2}{4\pi^2 I} \right)^2 \frac{N}{kT^2} \exp\left(-\frac{h^2}{4\pi^2 I k T}\right). \quad (44.12)$$

Thus, at very low temperatures the rotation energy and heat capacity turn out to decrease exponentially with the temperature. As we have stressed before, the decrease of the heat capacity with the temperature according to the exponential law can be observed only for the lightest molecules.

We write, in addition, the expressions for the free energy and entropy. At high temperatures

$$F_{\text{rot}} = -NkT \ln z_{\text{rot}} = -NkT \ln \left[\sum_{J=0}^{\infty} (2J+1) \exp \left(-\frac{T_c J(J+1)}{T} \right) \right] \approx \\ \approx NkT \left(\ln \frac{T_c}{T} - \frac{T_c}{3T} \right) \quad (44.13)$$

and

$$S_{\text{rot}} \approx -Nk \ln \frac{T_c}{T} + Nk. \quad (44.14)$$

At low temperatures, taking into account (44.7), we obtain

$$F_{\text{rot}} \approx -3NkT \exp(-2T_c/T) \quad (44.15)$$

and

$$S_{\text{rot}} \approx 3Nk \exp(-2T_c/T) + \frac{6NkT_c}{T} \exp(-2T_c/T). \quad (44.16)$$

Thus, the rotational energy and entropy decrease exponentially at very low temperatures. The general form of the heat capacity associated with rotation has the same character as that of the heat capacity associated with the vibrations of the molecules: at high temperatures the heat capacity tends to the classical value, while at low temperatures, in accordance with the requirements of the third law of thermodynamics, the heat capacity tends to zero. However, the notions of high and low temperature for rotations and vibrations turn out to be fundamentally different: for vibrations room temperature, as a rule, should be considered low, whereas for rotations it should be considered high.

As is seen from formulae (44.8), (44.12) and the subsequent formulae, it is necessary to know only one molecular constant, the moment of inertia I of the molecule, for the actual calculation of the partition function and thermodynamic quantities. The value of this quantity is known for most diatomic

molecules from spectroscopic data, in particular from infrared rotational spectra. If the rotational-vibrational spectrum of the molecule, i.e. the energy levels of its rotational and vibrational motions are known, then the calculation of the partition functions can be carried out by means of a direct summation *.

In conclusion we note that in the case of hydrogen and deuterium it is necessary to take into account the effect of the nuclear spin on the rotational motion. It turns out that the nuclear spin fundamentally affects the character of the rotational states of molecules consisting of identical atoms. In particular, depending on the value of the nuclear spin the molecules of hydrogen can be in rotational states of two types.

In states of the first type, corresponding to the total spin of the two nuclei being equal to zero, the rotational quantum number J runs over a series of even values $J = 0, 2, 4, 6, \dots$. In states of the second type, corresponding to the total spin of the two nuclei being equal to unity, the quantum number runs over a series of odd values $J = 1, 3, 5, \dots$.

Molecules of the first type are called *parahydrogen*, and those of the second type are called *orthohydrogen*. In normal conditions there are no transitions between orthohydrogen and parahydrogen, so that the gas as a whole must be considered as a mixture of the two different types.

This fact affects the form of the calculated thermodynamic functions ** in an essential way.

§45. Polyatomic molecules

The consideration of polyatomic molecules differs little in principle from that of diatomic molecules. The partition function of a polyatomic molecule, like that of a diatomic molecule, can be written in the form

$$Z = Z_{\text{transl}} Z_{\text{vib}} Z_{\text{rot}}, \quad (45.1)$$

provided the effect of the vibrations on the rotation of the molecule (in connection with the change in the size of the latter) is disregarded. The parti-

* For more detail see I.N.Godnev, *Vychislenie termodinamicheskikh funktsii po molekularnym dannym* (Calculation of thermodynamic functions from molecular data) (Gostekhizdat, Moscow, 1956). A.H.Wilson, *Thermodynamics and statistical mechanics* (Cambridge University Press, Cambridge, 1957).

** See, for example, L.D.Landau and E.M.Lifshitz, *Course in theoretical physics*, Volume 5, *Statistical physics* (Pergamon Press, London, 1958).

tion function of the translational motion differs in no way from that calculated before. However, the calculation of the partition function of the internal motion of polyatomic molecules is incomparably more complicated than for diatomic molecules.

In considering the rotational motion of a molecule it is necessary to distinguish between three cases: the linear molecule, the symmetric top, and the asymmetric top. The rotational motion of a linear polyatomic molecule does not differ from that of a diatomic molecule. For a symmetric top two major moments of inertia are equal to each other ($I_1 = I_2 \neq I_3$), whereas for an asymmetric top all moments of inertia differ from each other ($I_1 \neq I_2 \neq I_3$). In the first case quantum-mechanical considerations allow one to calculate the rotational energy levels of the molecule, which are expressed by a formula similar to that for the energy levels of a simple top. However, there is no explicit expression for the energy levels of an asymmetric top. In the case of molecules of the type of asymmetric tops use is usually made of certain approximate expressions for the energy levels, the accuracy of which is not very great. However, the situation is made substantially simpler by the fact that the characteristic temperatures for the rotation of polyatomic molecules are, as a rule, even lower than those for diatomic molecules. Hence the ordinary temperatures at which one can work with non-condensed polyatomic gases are relatively high, and for the partition function of the rotational motion use can be made, without any appreciable error, of the classical expression for z_{rot} . Thus, for example, the difference between the quantum and classical expression for the rotational state function for the HCN molecule at a temperature of 100 K amounts to about 0.5%, and for the CH_3Cl molecule it amounts to about 1%. At a temperature of 300 K this difference becomes quite negligible, and lies beyond the limits of accuracy of the measurements. In most calculations of the partition function of polyatomic molecules use is made of the classical approximation.

A characteristic feature of a large number of polyatomic molecules, in particular the molecules of organic compounds, is the presence of a certain number of identical atoms in them. The symmetry of a molecule is closely associated with the presence of identical atoms in the molecule. Because of the presence of symmetry the molecule can be matched with itself in definite rotations, just as a diatomic molecule containing two identical atoms matches with itself when rotated through 180° . The presence of symmetry in the molecule requires the introduction of a symmetry factor γ into the rotational partition function. A symmetry factor $\gamma = 2$ should also be introduced for diatomic molecules with identical nuclei. It represents the number of physically indistinguishable positions of the molecule when it rotates as a

rigid body. To obtain a correct expression for the rotational partition function, in which each physical state would be taken into account only once, the partition function obtained in integrating over all values of the angle of rotation (in the classical approximation) must be divided by the symmetry factor γ . With the introduction of the symmetry factor the rotational partition function of a polyatomic molecule with three different moments of inertia in the classical approximation can be written in the following form:

$$z_{\text{rot}} = \frac{\pi^{\frac{1}{2}}}{\gamma h^3} [8\pi^2(kT)^3 I_1 I_2 I_3]^{\frac{1}{2}}, \quad (45.2)$$

where the moments of inertia are expressed in $\text{g}\cdot\text{cm}^2$, and the temperature is measured on the absolute scale. If the structure of the molecule is known, then the values of the factor γ are found from simple considerations of the symmetry. In the case of the linear molecule CO_2 (see fig. III.19) the factor of symmetry is $\gamma = 2$, so that the molecule is matched with itself when rotated through the angle π . The non-linear molecule SO_2 (fig. III.19) is also matched with itself when rotated through the angle π , and so for it $\gamma = 2$. The methane molecule CH_4 represents a regular tetrahedron with the carbon atom in the centre. It is matched with itself when rotated through an angle of 120° about the vertical axis and when each of the four vertices of the tetrahedron are matched, all together in 12 rotations, so that $\gamma = 12$. The ammonia molecule NH_3 represents a pyramid with the nitrogen atom at the vertex. It is matched with itself when rotated through an angle of 120° about the vertical axis, so that $\gamma = 3$.

The vibrational motion of polyatomic molecules is also incomparably more complex than that of diatomic molecules. The number of vibrational degrees of freedom amounts to $3n - 6$ for non-linear polyatomic molecules and to $3n - 5$ for linear ones, and can be rather large for complex molecules. For example, the molecule SO_2 has three vibrational degrees of freedom, the molecule NH_3 already has six, and for the molecule C_6H_6 the number of vibrational degrees of freedom is equal to 30. The study of the vibrations of such systems is a complex problem. Nevertheless, the vibrational motion of a very large number of molecules has been investigated. If the departures of the atoms from the equilibrium position are assumed to be small (which is not always possible in the case of polyatomic molecules; see below), then the motion of the system will represent small vibrations and the vibrational motion of the molecule can be resolved into a set of independent normal

vibrations *. To each degree of freedom there corresponds one normal vibration with its natural frequency. The frequencies of the normal vibrations (the natural frequencies of the system) are connected with the masses of the nuclei and the constants of the quasi-elastic forces by the usual relations. In the general case the frequencies of all the normal vibrations are different. However, the frequencies of certain normal vibrations are often the same. In this case the vibrations are degenerate.

The values of the natural frequencies can be found from an analysis of infrared spectra as well as the dispersion spectra of the molecules, although this is not at all a simple problem. The spectra of a number of simpler molecules have been investigated in sufficient detail, and their natural frequencies have been determined with a high degree of accuracy. Fig. III.19 presents the normal vibrations of certain typical molecules (H_2O , SO_2 and CO_2). Each normal vibration with a natural frequency ν_l gives its contribution to the partition function z_{vib} which, by virtue of the independence of normal vibrations can be written in the form of the product of the corresponding factors:

$$z_{\text{vib}} = \prod_{l=1}^{3n-6} \frac{\exp(-h\nu_l/2kT)}{1 - \exp(-h\nu_l/kT)} = \prod_{l=1}^{3n-6} \frac{\exp(-T_c^{(l)}/2T)}{1 - \exp(-T_c^{(l)}/T)}, \quad (45.3)$$

where $T_c^{(l)}$ is the characteristic temperature of the l th normal vibration. The values of $T_c^{(l)}$ for different normal vibrations can differ considerably from each other. Thus, for example, the six normal oscillations of the ammonia molecule have the following characteristic temperatures $T_c^{(l)}$ (in 10^2 K): 13.6, 23.3, 23.3, 47.8, 48.8, 48.8. We see that characteristic temperatures can differ from each other by a factor of three. The difference between characteristic frequencies is associated with the difference between the constants of quasi-elastic forces (the difference in the rigidity of the atomic bonds in the molecules). The characteristic temperatures of polyatomic molecules, as well

* In arbitrary coordinates the potential energy of a system of vibrating points has the form

$$U = \frac{1}{2} \sum a_{ik} \xi_i \xi_k,$$

where ξ are the displacements. To find the normal vibrations it is necessary to find the coordinates ξ_i such that the potential energy of the system has the orthogonal quadratic form $U = \frac{1}{2} \sum a_j \xi_j^2$. The choice of new variables can be made in a purely algebraic way, but is simplified essentially if use is made of the symmetry properties of the system (see, for example, L.D. Landau and E.M. Lifshitz, *Mechanics* (Pergamon Press, Oxford, 1960).

as those of diatomic molecules, amount to several hundred or thousand degrees. Hence the contribution of vibrations to the heat capacity at moderate temperatures is relatively small. In any case, the vibrational part of the heat capacity is many times smaller than that to be expected from the law of equipartition.

As an example one can consider the ammonia molecule. The total heat capacity of the molecule NH_3 amounts to

$$C_V = C_{V_{\text{transl}}} + C_{V_{\text{rot}}} + C_{V_{\text{vib}}} = \\ = \frac{3}{2} + \frac{3}{2} + C_{V_{\text{vib}}} \quad (\text{in units of } Nk).$$

Table 7 gives the values (in $\text{J mol}^{-1} \text{ deg}^{-1}$) of the vibrational heat capacity $C_{V_{\text{vib}}}$ calculated according to (45.3) and observed experimentally.

Table 7

T (K)	$C_{V_{\text{vib}}}$ (calc.)	$C_{V_{\text{vib}}}$ (exper.)	T (K)	$C_{V_{\text{vib}}}$ (calc.)	$C_{V_{\text{vib}}}$ (exper.)
243	0.50	0.59	423	2.93	3.14
273	0.80	0.92	582	5.48	5.02
303	1.21	1.30	655	7.20	6.28
334	1.55	1.88	796	8.62	7.95
383	2.34	2.51			

For $T > 240 \text{ K}$ the rotational heat capacity can be assumed to have the classical value $\frac{3}{2}$. Thus, at a temperature of about 800 K the vibrational heat capacity amounts to about $\frac{2}{3}$ of the total heat capacity of the molecule and can in no way be disregarded. Even at room temperature $C_{V_{\text{vib}}}$ amounts to about 7% of the total heat capacity. Nevertheless, it is considerably smaller than that to be expected from the law of equipartition ($6Nk$). Such a situation is characteristic of most polyatomic molecules.

The contribution of different forms of motion to the value of the entropy of the ammonia molecule is shown in fig. III.28.

The vibrational motion of polyatomic molecules has a remarkable feature which has no analogue for diatomic molecules. Namely, very often the amplitude of zero-point vibrations of definite groups contained in the molecule turns out to be so large that the corresponding motion ceases to be harmonic or loses completely its vibrational character. This is seen most clearly by concrete examples. A large number of polyatomic molecules, in particular

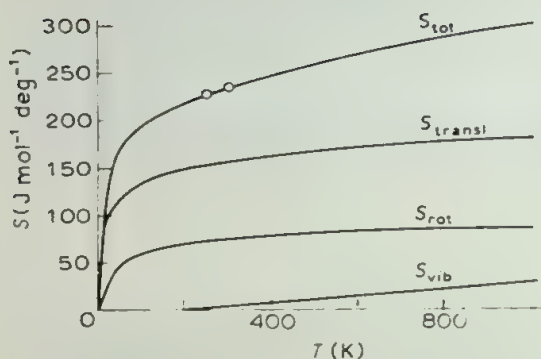


Fig. III.28

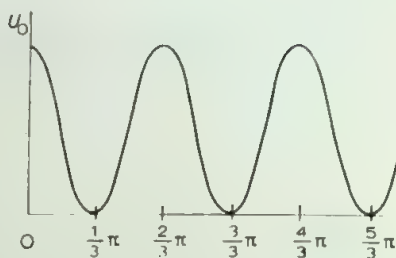


Fig. III.29

organic ones, contain individual groups or radicals having the character of independent groups; for example, the ethylene molecule C_2H_4 represents a combination of two CH_2 groups. Similarly the ethane molecule C_2H_6 consists of two CH_3 groups. The dimethylacetylene molecule $CH_3-C\equiv C-CH_3$ contains two CH_3 groups and a carbon core. Owing to the existence of interaction between the hydrogen atoms the potential energy of each of the CH_2 or CH_3 groups has a minimum for a definite orientation of one group with respect to the other. That is, to the minimum of the potential energy there corresponds a value of the angle of rotation (measured from the median line) of one CH_3 group with respect to the other equal to 60° or 180° . In other words, the potential energy has a minimum when the two groups are situated as an object and its mirror image (fig. III.29). When a displacement from the equilibrium position occurs (a rotation of one of the groups with respect to the other) the potential energy increases and there arises a force tending to bring the molecule back to the equilibrium arrangement. Then rotational oscillations about the axis of the molecule arise. If, however, the zero-point energy of these oscillations is so large that it exceeds the potential barrier hindering the group from rotation, the rotational oscillations turn over into a free rotation of the group with respect to the axis of the molecule. The latter case is encountered relatively seldom. The dimethylacetylene molecule quoted above serves as an example of it. In this molecule the two CH_3 groups are a relatively large distance apart, and their interaction is not very large. Hence the height of the barrier preventing a rotation turns out to be relatively small, and the CH_3 groups rotate freely. However, in most cases the rotation of individual groups is hindered. At low temperatures rotational oscillations

of a large amplitude take place, which turn over into a rotation at very high temperatures, when the thermal energy kT proves to be higher than the barrier height U_0 . The existence of free rotation changes the value of the heat capacity and other thermodynamic quantities in comparison with that for molecules without rotations: a part of the vibrational degrees of freedom is replaced by rotational ones. If the rotation is assumed to be free, then the calculation of the heat capacity and other thermodynamic quantities presents no difficulty, since the rotation of a relatively heavy group can be assumed to be classical. But if the rotation is hindered, then it is necessary to know for the calculation the height of the barrier hindering the rotation. Finding this height from spectroscopic data is very difficult. Hence one proceeds in the reverse order: one calculates thermodynamic quantities, most often the entropy, by assuming different values of the height of the barrier, and compares the calculated and measured values. The matching of the theoretical and experimental curve of the dependence of the entropy on the temperature allows one to find the height of the barrier. For different molecules the height of the barrier varies within a rather wide range. Thus, for the rotation of the CH_3 group in the ethane molecule $\text{H}_3\text{C}-\text{CH}_3$ the height of the barrier amounts to about 1570 K, so that the rotation at a temperature $T > 1570$ K is free. For the ethylene molecule C_2H_4 the barrier hindering the CH_2 groups from rotation has a height of about 6000 K, so that the rotation at room temperature is strongly hindered and there take place rotational oscillations of a relatively small amplitude.

Systems of Interacting Particles

§46. Interaction between molecules in non-ideal gases

Up to now we have restricted ourselves to the study of the properties of gases so rarefied that the interaction between the molecules may be disregarded. We now pass on to the consideration of the statistical behaviour of systems of interacting particles.

In §6 we have touched already upon the problem of the character of the intermolecular interaction. At large distances between molecules this interaction reduces to a weak attractive force which decreases rapidly with increasing distance between the centres of the molecules. At small distances, when molecules approach each other closely, so that a mutual penetration of their electron shells takes place, a very strong repulsion arises. Owing to this repulsion the molecules cannot penetrate very much into each other, and they cannot get deformed in collisions. In what follows we shall confine ourselves to a monatomic gas, and shall assume that the interaction depends only on the distance between the atoms.

The form of the potential energy of the interaction between two molecules is shown in fig. III.30. We shall assume that the attractive forces are so weak that the largest value of the potential energy of attraction occurs when the molecules approach closely (the distance between the centres is

equal to the diameter d). But $|u(d)|$ is nevertheless small in comparison with the thermal energy kT ,

$$|u(d)| \ll kT. \quad (46.1)$$

The potential energy decreases so rapidly with increasing distance that it is practically reduced to zero at distances between the centres of the molecules amounting to only a few diameters. We formally introduce a certain distance ρ such that beyond it the interaction can be completely disregarded. We shall call this distance the radius of interaction. This means that we replace the true potential energy curve by the simplified curve shown by a dotted line in fig. III.30 and expressed by the formulae

$$u = \begin{cases} 0, & r \geq \rho, \\ -u(r), & \rho > r > d, \\ \infty, & d \geq r. \end{cases} \quad (46.2)$$

Here r denotes the distance between the centres of the i th and k th molecules. This means that the interaction is absent when the distance between the centres of the molecules exceeds ρ , represents an attraction [negative value of $u(r)$] when the distance between the centres is smaller than ρ but larger than d , so that the molecules do not directly touch each other, and turns into a very strong repulsion (infinitely strong in our approximation)

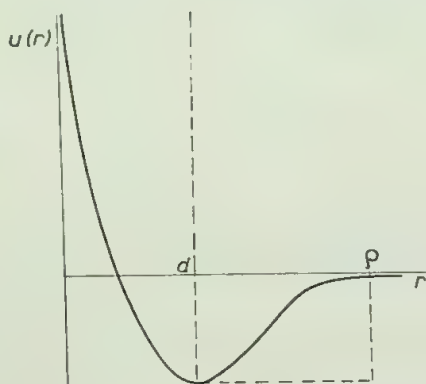


Fig. III.30

when the molecules come into direct contact. The value of ρ is as a rule equal to three or four molecular diameters.

If the gas is not very dense, then the mean distance between the molecules is very large in comparison with their dimensions. Hence it can as a rule be assumed that not more than two molecules approach each other to within the distance of interaction at once. In other words, it can be assumed that the molecules interact only in pairs. Configurations in which a "cluster" of three, four, or more particles are found simultaneously in the sphere of interaction are seldom encountered, and we shall disregard them.

We calculate the partition function of the gas in this condition. The energy ϵ of the entire gas can be written in the form

$$\epsilon = \epsilon_{\text{kin}} + U. \quad (46.3)$$

The first term in (46.3) expresses the sum of the kinetic energies of the molecules. It is the same as the energy of an ideal gas. The second term represents the potential energy of interaction of the molecules, depending only on their mutual distances. Making use of this expression for ϵ , one can write the partition function of the gas in the form

$$Z = \frac{1}{N!h^{3N}} \left(\int e^{-p^2/2mkT} dp_x dp_y dp_z \right)^N \left(\int e^{-U/kT} dV_1 \dots dV_N \right), \quad (46.4)$$

where $dV_1 \dots dV_N$ is the product of the differentials of the space coordinates $dx dy dz$ for every one of the molecules. The first factor does not differ from the corresponding quantity for an ideal gas. By virtue of the results of §37 it can be written in the form

$$\int e^{-p^2/2mkT} dp_x dp_y dp_z = (2\pi mkT)^{\frac{3}{2}},$$

so that

$$Z = \frac{1}{N!h^{3N}} (2\pi mkT)^{\frac{3}{2}N} \left(\int e^{-U/kT} dV_1 \dots dV_N \right). \quad (46.5)$$

Here we have to calculate the second factor, called the configuration integral:

$$I = \int e^{-U/kT} dV_1 dV_2 \dots dV_N. \quad (46.6)$$

To do this we make use of the assumption that the molecules interact with one another only in pairs. Hence the energy of interaction can be written in the form of the sum of the energies of interaction of the molecular pairs

$$U = \sum u(r_{ik}), \quad (46.7)$$

where a pair is understood to be two molecules which are at a distance smaller than the radius of interaction ρ . The energy of interaction of each pair is denoted by u . It is determined by formula (46.2). The number of terms in the sum (46.7) is equal to the number of pairs formed in a gas of N molecules. It is equal to the number of combinations of N elements in twos, i.e. to $\frac{1}{2}N(N-1)$. For large N this number can be assumed to be equal to $\frac{1}{2}N^2$. Then

$$\exp\left(-\frac{U}{kT}\right) = \exp\left(-\frac{\sum u(r_{ik})}{kT}\right) = \prod_i \exp\left(-\frac{u(r_{ik})}{kT}\right),$$

where the product is taken over all the pairs, i.e.

$$\exp\left(-\frac{U}{kT}\right) = \exp\left(-\frac{u(r_{12})}{kT}\right) \exp\left(-\frac{u(r_{13})}{kT}\right) \dots \quad (46.8)$$

This product contains $\frac{1}{2}N^2$ factors. Each term in this product tends to unity when $r_{ik} > \rho$, since $u(r_{ik}) \rightarrow 0$. It is more convenient to introduce a function f_{ik} defined by

$$f_{ik} = \exp\left(-\frac{u(r_{ik})}{kT}\right) - 1, \quad (46.9)$$

which tends to zero when $r_{ik} > \rho$ and differs from zero only when $r_{ik} < \rho$. Then, obviously, $e^{-u(r_{ik})/kT} = 1 + f_{ik}$, and

$$\begin{aligned} e^{-U/kT} &= \prod (1+f_{ik}) = \\ &= (1+f_{12})(1+f_{13})(1+f_{14}) \dots = \\ &= 1 + (f_{12}+f_{13}+f_{14}+\dots) \\ &\quad + (f_{12}f_{13}+f_{12}f_{14}+\dots) + \dots \end{aligned} \quad (46.10)$$

Indeed, the double, triple and so on products of the function f_{ik} are, by definition of this function and by virtue of the assumption of the absence of clusters, always very small. Thus, for example, in order that $f_{12}f_{13}$ may differ substantially from zero it is necessary that f_{12} and f_{13} should at the same time be different from zero, i.e. that the distances r_{12} and r_{13} should at the same time be small (smaller than ρ). This means that the first, second and third molecules simultaneously come into the range of interaction ρ , forming not a pair but a triplet of molecules. Similarly $f_{12}f_{13}f_{14}$ differs from zero only if f_{12} , f_{13} and f_{14} are not at the same time equal to zero. This happens only when the first, second, third and fourth molecules simultaneously get into a region of the order of ρ . Hence with a sufficient degree of accuracy it can be written that

$$e^{-U/kT} \approx 1 + (f_{12} + f_{13} + \dots) = 1 + \sum f_{ik}. \quad (46.11)$$

The number of terms in $\sum f_{ik}$ is equal to the number of pairs, i.e. to $\frac{1}{2}N^2$. Since all the molecules are identical, it can be assumed that all f_{ik} are the same, so that

$$e^{-U/kT} \approx 1 + \frac{1}{2}N^2 f(r_{ik}). \quad (46.12)$$

Substituting the expression $e^{-U/kT}$ from (46.12) into (46.6), we have

$$\begin{aligned} I &= \int e^{-U/kT} dV_1 \dots dV_N = \int (1 + \frac{1}{2}N^2 f_{ik}) dV_1 \dots dV_N = \\ &= \int dV_1 \dots dV_N + \frac{1}{2}N^2 \int f_{ik} dV_1 \dots dV_N. \end{aligned} \quad (46.13)$$

The first integral in (46.13) is obviously equal to V^N . In the second integral the integration over all volume elements except the i th and k th gives

$$\begin{aligned} \int dV_1 \dots dV_{i-1} dV_{i+1} \dots dV_{k-1} dV_{k+1} \dots dV_N \int f_{ik} dV_i dV_k = \\ = V^{N-2} \int f_{ik} dV_i dV_k. \end{aligned}$$

Thus,

$$I = V^N + \frac{1}{2}N^2 V^{N-2} \int f_{ik} dV_i dV_k.$$

To carry out the last integration we introduce spherical coordinates with the centre located at one of the molecules. Then $r_{ik} = r$, and

$$\begin{aligned}\int f_{ik} dV_i dV_k &= \int \left[\exp \left(-\frac{u(r_{ik})}{kT} \right) - 1 \right] dV_i dV_k = \\ &= \int dV \int \left[\exp \left(\frac{u(r)}{kT} \right) - 1 \right] 4\pi r^2 dr ,\end{aligned}$$

where 4π is the result of integration with respect to the angles. Hence, writing

$$4\pi \int \left[\exp \left(-\frac{u(r)}{kT} \right) - 1 \right] r^2 dr = \beta , \quad (46.14)$$

we obtain

$$\int f_{ik} dV_i dV_k = V\beta .$$

For I the final expression is the following:

$$I = V^N \left(1 + \frac{N^2}{2V} \beta \right) . \quad (46.15)$$

Substituting the expression (46.15) for I into (46.5), we have

$$Z = \frac{(2\pi mkT)^{\frac{3}{2}N} V^N}{N! h^{3N}} \left(1 + \frac{N^2 \beta}{2V} \right) = Z_{\text{id}} \left(1 + \frac{N^2 \beta}{2V} \right) , \quad (46.16)$$

where Z_{id} denotes the partition function of the ideal monatomic gas. It should be noted that the value of $N^2 \beta / 2V = NV^{-1} \cdot \frac{1}{2} N \beta$ is small for a small gas density NV^{-1} .

By means of the partition function (46.16) one can calculate the thermodynamic functions of a gas which departs slightly from an ideal gas. We shall restrict ourselves to the calculation of the pressure, since the equation of state of a gas is of very great interest.

The departure of a gas from an ideal gas is taken into account by means of the van der Waals equation, which for small gas densities can be written in the form

$$p = \frac{NkT}{V - Nb} - \frac{N^2 a}{V^2} \approx \frac{NkT}{V} + \frac{N^2 kTb}{V^2} - \frac{N^2 a}{V^2} . \quad (46.17)$$

Since in the course of the calculations of the preceding paragraph we have not taken into account "clusters", the results obtained hold for small gas densities. A simple calculation of the pressure, based on the partition function (46.16), leads to an expression which is exactly the same as (46.17). Indeed, according to (32.5) the pressure p is equal to

$$\begin{aligned} p &= kT \frac{\partial \ln Z}{\partial V} = kT \frac{\partial \ln Z_{id}}{\partial V} + kT \frac{\partial}{\partial V} \ln \left(1 + \frac{N^2 \beta}{2V} \right) \approx \\ &\approx p_{id} + kT \frac{\partial}{\partial V} \left(\frac{N^2 \beta}{2V} \right) = \frac{NkT}{V} - \frac{N^2 kT \beta}{2V^2}. \end{aligned} \quad (46.18)$$

Here, assuming the density to be small, we have expanded the logarithm in a series in powers of the quantity $N^2 \beta / 2V$ which is very small in comparison with unity, and have confined ourselves to the first term of the expansion.

Comparing formula (46.18) with (46.17) we convince ourselves of their complete identity provided we assume

$$\frac{1}{2} \beta = \frac{a}{kT} - b. \quad (46.19)$$

Thus, formula (46.18) represents the van der Waals equation, derived theoretically for small gas densities. The foregoing calculation refers to the case of monatomic gases. It can, however, also be shown that in the case of complex polyatomic gases the qualitative aspect of the derivation is not changed, although the explicit form of the quantity β will be more complicated.

To explain the meaning of the constants a and b occurring in the van der Waals equation, we shall consider in detail the quantity β .

By definition,

$$\beta = 4\pi \int_0^\infty f(r) r^2 dr. \quad (46.20)$$

Substituting the expression for $f(r)$ into (46.20), we have

$$\beta = 4\pi \int_0^\infty \left[\exp \left(-\frac{u(r)}{kT} \right) - 1 \right] r^2 dr. \quad (46.21)$$

We divide the integral (46.21) into two parts: the integral in the interval $0 \leq r \leq d$ and the integral in the interval $d \leq r \leq \infty$, i.e.

$$\beta = 4\pi \int_0^d \left[\exp\left(-\frac{u(r)}{kT}\right) - 1 \right] r^2 dr + 4\pi \int_d^\infty \left[\exp\left(-\frac{u(r)}{kT}\right) - 1 \right] r^2 dr.$$

In the first interval, by virtue of (46.2), $e^{-u(r)/kT} \approx 0$, and thus in the first integral the exponential term can be dropped. In the second integral the potential energy of interaction of the molecules is, by virtue of (46.1), small in comparison with the thermal energy kT , so that it can be written approximately that

$$\exp\left(-\frac{u(r)}{kT}\right) \approx 1 - \frac{u(r)}{kT} = 1 + \frac{|u(r)|}{kT}.$$

We then have

$$\beta = -4\pi \int_0^d r^2 dr + \frac{4\pi}{kT} \int_d^\infty |u(r)| r^2 dr = -\frac{4\pi}{3} d^3 + \frac{4\pi}{kT} \int_d^\infty |u(r)| r^2 dr.$$

Substituting β into (46.19), we obtain

$$-\frac{2\pi d^3}{3} + \frac{2\pi}{kT} \int_d^\infty |u(r)| r^2 dr = \frac{a}{kT} - b. \quad (46.22)$$

Equating in (46.22) the coefficients of T^{-1} and the constant terms, we find

$$b = \frac{2}{3}\pi d^3 = 4v_0, \quad (46.23)$$

where v_0 is the volume occupied by the molecule. Thus, the constant b in the van der Waals equation turns out to be equal to the volume of the molecule multiplied by four. Further,

$$a = \frac{1}{2} \cdot 4\pi \int_d^\infty |u(r)| r^2 dr.$$

The constant a is expressed in terms of the integral of the potential energy of interaction of two molecules. Since the function $u(r)$ rapidly decreases with increasing distance between the molecules, this integral converges rapidly. Thus,

$$\beta = \frac{2a}{kT} - 8v_0. \quad (46.24)$$

Depending on the temperature, β can be positive or negative. At a sufficiently low temperature $\beta > 0$, while at a high temperature $\beta < 0$.

If the expressions which we have found for the constants a and b are substituted into the van der Waals equation, we then obtain

$$p = \frac{NkT}{V} \left(1 + \frac{4v_0N}{V} - \frac{2\pi N}{VkT} \int_d^\infty |u(r)|r^2 dr \right). \quad (46.25)$$

In the first approximation, when the gas density is so low that the probability of the simultaneous presence of three or more molecules in the sphere of interaction may be disregarded, the pressure in a non-ideal gas differs from that in an ideal gas by two terms. The first of these represents the ratio of four times the volume of all the molecules to the entire volume of the gas. The meaning of this (positive) correction to the pressure lies in the fact that it takes into account the volume of real molecules. In this approximation we cannot consider molecules as material points having no spatial extension.

The second correction to the pressure is negative and is in absolute value equal to the ratio

$$\frac{N^2}{2V} \left(\frac{4\pi}{V} \int_d^\infty |u(r)|r^2 dr \right).$$

This ratio also has a simple physical meaning. The quantity

$$\frac{1}{V} 4\pi \int_d^\infty |u(r)|r^2 dr = \bar{u}$$

represents the mean value of the potential energy of interaction of a pair of molecules. This mean value is taken over all possible distances between the molecules, i.e. over the entire volume available for the motion of the molecules. Then, obviously, $N^2\bar{u}/2V$ is the mean value of the energy of interaction of all pairs of molecules existing in a unit volume of the gas. The second correction thus characterizes the decrease in the pressure of gas molecules on the walls of the container due to their mutual attraction. This fact can be expressed in another way as follows: in the gas there exists an internal pressure due to the attraction of molecules.

As is well known, the van der Waals equation describes not only the properties of gases with a relatively low density but also those of very dense gases and even liquids. However, in this case it cannot be derived theoretically, and represents a purely empirical equation which should be considered as a

more or less successful extrapolation from the region of low densities. How this extrapolation is to be made is seen from eq. (46.17), if it is rewritten in the form

$$p = \frac{NkT}{V} \left(1 + \frac{Nb}{V} \right) - \frac{N^2a}{V^2}. \quad (46.26)$$

The equation is valid for $Nb/V \ll 1$, i.e. when the volume $4Nv_0$ occupied by all the molecules is very small in comparison with the volume of the gas.

If, however, the density of the gas increases, which can be characterized by a decrease in the volume V for a fixed N , then formula (46.26) loses its validity. It is physically clear that, if the gas is compressed to the limit of close packing of the molecules with the minimum gap between them, a correct formula for the pressure should point to an infinite increase in it. A further compression, associated with a deformation of the atoms, would be associated with enormous pressures which would be infinitely large in comparison with ordinary pressures in gases or liquids. From geometrical considerations it is clear that to the close packing of spherical molecules there corresponds a volume of the system equal to $4v_0N = Nb$. Consequently, a correct formula for the pressure should lead to indefinitely increasing values of p as $V \rightarrow Nb$. However, formula (46.26) does not have such a character. But if the factor $1 + NbV^{-1}$ is considered as a result of the expansion in a series of the quantity $(1 - NbV^{-1})^{-1}$, then we obtain immediately the following formula for the pressure:

$$p = \frac{NkT}{V(1 - NbV^{-1})} - \frac{N^2a}{V^2}, \quad (46.27)$$

which satisfies the required conditions:

- 1) p increases to infinity as $V \rightarrow Nb$,
- 2) for $V \gg Nb$ formula (46.27) goes over into the theoretical formula (46.17).

Formula (46.27) is the complete van der Waals equation which describes the state of gases over a wide range of densities. From the very character of the derivation it is clear, however, that this equation cannot have the important theoretical meaning which is possessed by eq. (46.17). For large gas densities the constants a and b no longer have an exact meaning, and can only approximately be considered as the characteristics of the volume of the molecules and their interaction. This is seen, in particular, from the fact that in order to obtain quantitative agreement of eq. (46.27) with experimental data one has to renounce the constancy of the quantities a and b and con-

sider them as functions of temperature. The inconvenience of this led many investigators to propose other empirical equations of state. Nevertheless, a great merit of the van der Waals equation is the fact that it gives qualitatively, in a very correct way, the behaviour of gases, and it contains indications about the transition of a gas into the liquid state and critical phenomena.

§47. The correlation function method and its application to the theory of dense gases and liquids

We have seen that a direct calculation of the configuration integral for a system of interacting particles turns out to be a rather complex procedure.

In connection with the attempt to create a statistical theory of liquids, different methods of approach to the consideration of the statistical properties of systems of interacting particles have recently been developed. The correlation function method developed by N.N. Bugoliubov and independently by Kirkwood and by Born and Green has proved to be one of the most effective.

In the correlation function method, the calculation of the configuration integral is replaced by finding a certain set of integro-differential equations interrelating a system of functions which characterize the mutual correlation in the spatial disposition of the particles.

We stress from the very beginning that the correlation function method is a direct consequence of Gibbs statistics. In what follows we shall be interested in the spatial distribution of a system of interacting particles.

Integrating the Gibbs distribution over all momenta, we find the expression for the probability of a given configuration of a system of particles:

$$dw_{\mathbf{r}} = \int_{\mathbf{p}} dw = I^{-1} e^{-U/kT} d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_N, \quad (47.1)$$

where I is the configuration integral:

$$I = \int e^{-U/kT} d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_N. \quad (47.2)$$

If $dw_{\mathbf{r}}$ is integrated over the coordinates of all the particles except one, then we obtain

$$dw_{\mathbf{r}}^{(1)} = \frac{d\mathbf{r}_1}{I} \int e^{-U/kT} d\mathbf{r}_2 \dots d\mathbf{r}_N. \quad (47.3)$$

It is obvious that $dw_{\mathbf{r}}^{(1)}$ represents the probability that particle no. 1 be in the volume element $d\mathbf{r}_1$ for any positions of all the other $N-1$ particles. This probability can be written in the form

$$dw_{\mathbf{r}}^{(1)} = \frac{\rho_1(\mathbf{r}_1)d\mathbf{r}_1}{V}, \quad (47.4)$$

where $\rho_1(\mathbf{r}_1)$ is the density of the probability that the particle be found in the volume element $d\mathbf{r}_1$ normalized to the volume of the system:

$$\frac{\rho_1(\mathbf{r}_1)}{V} = \frac{1}{I} \int e^{-U/kT} d\mathbf{r}_2 \dots d\mathbf{r}_N. \quad (47.5)$$

We shall call the function $\rho_1(\mathbf{r}_1)$ the ordinary distribution function.

Analogously, integrating the Gibbs distribution (47.1) over the coordinates of all the particles except the first and second, we obtain

$$dw_{\mathbf{r}}^{(1,2)} = \frac{d\mathbf{r}_1 d\mathbf{r}_2}{I} \int e^{-U/kT} d\mathbf{r}_3 \dots d\mathbf{r}_N = \frac{\rho_{12}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2}{V^2}, \quad (47.6)$$

so that

$$\frac{\rho_{12}(\mathbf{r}_1, \mathbf{r}_2)}{V^2} = \frac{\rho_{12}}{V^2} = \frac{1}{I} \int e^{-U/kT} d\mathbf{r}_3 \dots d\mathbf{r}_N. \quad (47.7)$$

The function ρ_{12} represents the density of the probability that the first particle is in the volume element $d\mathbf{r}_1$ and that the second particle is simultaneously in the volume element $d\mathbf{r}_2$ normalized to the volume of the system.

We shall call ρ_{12} the binary distribution function. Distribution functions of any order can be determined in an analogous way. For example, the distribution function of the m th order characterizes the probability that the first particle be in the volume element $d\mathbf{r}_1$, the second particle in the volume element $d\mathbf{r}_2$, the m th particle in the volume element $d\mathbf{r}_m$ for any positions of the other $N-m$ particles:

$$\frac{\rho_{12\dots m}(\mathbf{r}_1, \dots, \mathbf{r}_m)}{V^m} = \frac{1}{I} \int e^{-U/kT} d\mathbf{r}_{m+1} \dots d\mathbf{r}_N. \quad (47.8)$$

If we are interested in the properties of a system which depend on the position not of all but only several particles constituting the system, the distribu-

tion functions $\rho_{12\dots m}$ play the same role as the Gibbs distribution function for the system as a whole.

By means of distribution functions one can find the mean values of quantities which depend on the coordinates of the corresponding particles. For example

$$L(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = \int L(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) \rho_{12\dots m}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) \frac{d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_n}{V^m}.$$

§48. Equations for correlation functions

At first sight it may seem that finding the distribution function of m th order which characterizes the spatial distribution of certain particles of a system must be simpler than finding the Gibbs distribution, the distribution function of N th order characterizing the configuration of all the particles of the system.

However, it is clear that a direct determination of the distribution functions $\rho_1, \rho_2, \dots, \rho_m, \dots$ is associated with the calculation of the configuration integral and hence their application in no way simplifies the problem and is no step forward.

The application of the distribution functions would be of no interest if there were not another method of calculating them which is not associated with the determination of I .

It turns out to be possible to obtain a differential equation which must be satisfied by the distribution functions ρ_m . To find an equation which must be satisfied by an ordinary function, we differentiate formula (47.5) with respect to the coordinates \mathbf{r}_1 . Obviously, we have

$$\frac{\partial \rho_1(\mathbf{r}_1)}{\partial \mathbf{r}_1} = - \frac{V}{IkT} \int e^{-U/kT} \frac{\partial U}{\partial \mathbf{r}_1} d\mathbf{r}_2 \dots d\mathbf{r}_N. \quad (48.1)$$

Consider in more detail the derivative

$$\frac{\partial U}{\partial \mathbf{r}_1} = \frac{\partial}{\partial \mathbf{r}_1} \sum_{1 \leq i < j \leq N} u(|\mathbf{r}_i - \mathbf{r}_j|) = \frac{\partial}{\partial \mathbf{r}_1} \sum_{j=2}^N u(|\mathbf{r}_1 - \mathbf{r}_j|), \quad (48.2)$$

[cf. eq. (46.7)]. Here we have made use of the fact that all terms of the sum over i except the one referring to particle no. 1 do not depend on \mathbf{r}_1 and reduce to zero in differentiating.

Substituting (48.2) into (48.1), we obtain on the right-hand side the integral

$$\begin{aligned}
 -\frac{V}{IkT} \int e^{-U/kT} \frac{\partial}{\partial \mathbf{r}_1} \sum_j u(|\mathbf{r}_1 - \mathbf{r}_j|) d\mathbf{r}_2 \dots d\mathbf{r}_N = \\
 = -\frac{V}{IkT} \sum_j \int e^{-U/kT} \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_j|)}{\partial \mathbf{r}_1} d\mathbf{r}_2 \dots d\mathbf{r}_N.
 \end{aligned}$$

But, by the definition (47.8),

$$\frac{1}{I} \int e^{-U/kT} d\mathbf{r}_2 \dots d\mathbf{r}_{j-1} d\mathbf{r}_{j+1} \dots d\mathbf{r}_N = \frac{\rho_{1j}(\mathbf{r}_1, \mathbf{r}_j)}{V^2}.$$

Hence the right-hand side can be written in the form

$$\frac{1}{VkT} \sum_{j=2}^N \int \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_j|)}{\partial \mathbf{r}_1} \rho_{1j}(\mathbf{r}_1, \mathbf{r}_j) d\mathbf{r}_j.$$

The sum over j contains $N - 1$ terms each of which represents an integral of the form

$$\int \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_j|)}{\partial \mathbf{r}_1} \rho_{1j}(\mathbf{r}_1, \mathbf{r}_j) d\mathbf{r}_j.$$

Since the system consists of identical particles, so that $u(|\mathbf{r}_1 - \mathbf{r}_j|)$ for given $|\mathbf{r}_1 - \mathbf{r}_j|$ has one and the same value and the integration is carried out over all $|\mathbf{r}_1 - \mathbf{r}_j|$, one can write

$$\begin{aligned}
 \frac{1}{V} \sum_{j=2}^N \int \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_j|)}{\partial \mathbf{r}_1} \rho_{1j}(\mathbf{r}_1, \mathbf{r}_j) d\mathbf{r}_j = \\
 = \frac{N-1}{V} \int \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_j|)}{\partial \mathbf{r}_1} \rho_{1j} d\mathbf{r}_j \approx \\
 \approx \frac{N}{V} \int \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_j|)}{\partial \mathbf{r}_1} \rho_{1j} d\mathbf{r}_j.
 \end{aligned}$$

Substituting this expression into formula (48.1), we obtain

$$\frac{\partial \rho_1(\mathbf{r}_1)}{\partial \mathbf{r}_1} = -\frac{N}{VkT} \int \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_j|)}{\partial \mathbf{r}_1} \rho_{1j} d\mathbf{r}_j. \quad (48.3)$$

Formula (48.3) relates the ordinary distribution function ρ_1 to the binary distribution function ρ_{1j} .

Let us find the equation satisfied by the binary function ρ_{1j} . Differentiating (47.7), we obtain

$$\begin{aligned} \frac{\partial \rho_{12}(\mathbf{r}_1, \mathbf{r}_2)}{\partial \mathbf{r}_1} &= -\frac{V^2}{IkT} \int e^{-U/kT} \frac{\partial}{\partial \mathbf{r}_1} \left[\sum u(|\mathbf{r}_1 - \mathbf{r}_j|) \right] d\mathbf{r}_3 \dots d\mathbf{r}_N = \\ &= -\frac{V^2}{IkT} \int e^{-U/kT} \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_2|)}{\partial \mathbf{r}_1} d\mathbf{r}_3 \dots d\mathbf{r}_N \\ &\quad - \frac{V^2}{IkT} \sum_{j=3}^N \int \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_j|)}{\partial \mathbf{r}_1} d\mathbf{r}_j \int e^{-U/kT} d\mathbf{r}_3 \dots d\mathbf{r}_{j-1} d\mathbf{r}_{j+1} \dots d\mathbf{r}_N \approx \\ &\approx -\frac{N}{VkT} \int \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_j|)}{\partial \mathbf{r}_1} \rho_{12j} d\mathbf{r}_j - \frac{1}{kT} \rho_{12}(\mathbf{r}_1, \mathbf{r}_2) \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_2|)}{\partial \mathbf{r}_1}. \end{aligned}$$

Consequently, we have finally

$$\frac{\partial \rho_{12}}{\partial \mathbf{r}_1} = -\frac{1}{kT} \rho_{12} \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_2|)}{\partial \mathbf{r}_1} - \frac{N}{VkT} \int \frac{\partial u(|\mathbf{r}_1 - \mathbf{r}_j|)}{\partial \mathbf{r}_1} \rho_{12j} d\mathbf{r}_j. \quad (48.4)$$

Eq. (48.4) connects the binary distribution function with the threefold distribution function. In the same way, one can obtain an equation connecting the threefold distribution function with the fourfold, the m th distribution function with the $(m+1)$ th, and so on. As a result one obtains an unclosed system of equations each of which expresses the derivative of the distribution function of a given order in terms of the distribution function of the next order:

$$\begin{aligned} \frac{\partial \rho_{12\dots m}(\mathbf{r}_1, \dots, \mathbf{r}_m)}{\partial \mathbf{r}_1} &= -\frac{1}{kT} \rho_{12\dots m} \frac{\partial}{\partial \mathbf{r}_1} \left[\sum u(|\mathbf{r}_1 - \mathbf{r}_m|) \right] \\ &\quad - \frac{N}{VkT} \int \sum_{j=m+1}^N \frac{\partial u}{\partial \mathbf{r}_1} \rho_{12\dots m, m+1} d\mathbf{r}_{m+1}. \end{aligned} \quad (48.5)$$

Continuing this procedure we shall arrive at an equation connecting the distribution function of the $(N-1)$ th order with the distribution function

of the N th order, i.e. with the Gibbs distribution. Hence the problem of finding distribution functions of a lower order again turns out to be associated with the Gibbs distribution for the entire system. However, in this fact lies the most important feature of the equations obtained: functions of a higher order always enter under the integral sign with a coefficient $\sim (N/kT)(\partial u/\partial \mathbf{r}_1)$. In the case where the potential energy of interaction between two particles, $u(|\mathbf{r}_1 - \mathbf{r}_j|)$, decreases rapidly with increasing distance and becomes small at distances which exceed molecular dimensions the quantity $\partial u/\partial \mathbf{r}_1$ is very small for $|\mathbf{r}_1 - \mathbf{r}_j| \gg d$, where d is the diameter of the molecule. Hence, for example, the expression for the integral in the right-hand side of eq. (48.4) can be estimated to order of magnitude in the following way:

$$\frac{N}{V} \int \frac{\partial u}{\partial \mathbf{r}_1} \rho_{12j} d\mathbf{r}_j \approx \frac{d^3}{V/N} \left[\frac{\partial u}{\partial \mathbf{r}_1} \rho_{12j} \right]_d,$$

where $[(\partial u/\partial \mathbf{r}_1) \rho_{12j}]_d$ is taken for distances between the particles of the order of d .

If the volume per particle, V/N , is large in comparison with the volume d^3 of the particle, then the coefficient $d^3/(VN^{-1})$ is small. Hence use can be made of approximate expressions for the value of the integrand, in particular the distribution function of the third order. This conclusion refers not only to the equation for the binary function, but has a general character.

§49. The equation of state and the energy of a system of interacting particles

The binary distribution function $\rho_{12}(\mathbf{r}_1, \mathbf{r}_2)$, in terms of which the equation of state can be expressed is of great importance.

The pressure in a system is determined by formula (32.5):

$$p = \frac{kT}{Z} \frac{\partial Z}{\partial V}.$$

Since, as we have seen before, Z resolves into two factors, one of which, Z_{kin} , depends only on the kinetic energy but not on the configuration of the particles and, consequently, not on the volume, and the second of which, I , depends only on the configuration, we have

$$p = \frac{kT}{Z_{\text{kin}} I} \frac{\partial Z_{\text{kin}} I}{\partial V} = \frac{kT}{I} \frac{\partial I}{\partial V}. \quad (49.1)$$

We shall find the derivative of I with respect to the volume of the system. For this we change all linear dimensions by a factor λ :

$$\mathbf{r}^* \rightarrow \lambda \mathbf{r}. \quad (49.2)$$

In consequence of (49.2) we obtain

$$V^* \rightarrow \lambda^3 V. \quad (49.3)$$

Then

$$I = \lambda^{3N} \int e^{-U/kT} d\mathbf{r}_1 \dots d\mathbf{r}_N. \quad (49.4)$$

According to (49.3)

$$\frac{\partial Z}{\partial V^*} = \frac{\partial I}{\partial V^*} = \frac{\partial I}{\partial \lambda} \frac{\partial \lambda}{\partial V^*} = \frac{\partial I}{\partial \lambda} \frac{1}{3\lambda^2 V}$$

and, consequently,

$$\frac{\partial I}{\partial V^*} = \frac{1}{3V} \left. \frac{\partial I}{\partial \lambda} \right|_{\lambda=1}. \quad (49.5)$$

Differentiating (49.4) with respect to λ , we obtain

$$\frac{\partial I}{\partial \lambda} = \frac{3NI}{\lambda} - \frac{\lambda^{3N}}{kT} \int e^{-U/kT} \frac{\partial U}{\partial \lambda} d\mathbf{r}_1 \dots d\mathbf{r}_N. \quad (49.6)$$

According to the definition (46.7)

$$\frac{\partial U}{\partial \lambda} = \frac{\partial}{\partial \lambda} \sum \sum u(\lambda |\mathbf{r}_i - \mathbf{r}_j|) = \sum \sum (|\mathbf{r}_i - \mathbf{r}_j|) u'.$$

Substituting $\partial U / \partial \lambda$ into (49.6), we find

$$\left. \frac{\partial I}{\partial \lambda} \right|_{\lambda=1} = 3NI - \frac{1}{kT} \sum \sum \int (|\mathbf{r}_i - \mathbf{r}_j|) u' e^{-U/kT} d\mathbf{r}_1 \dots d\mathbf{r}_N,$$

or

$$\left. \frac{\partial I}{\partial \lambda} \right|_{\lambda=1} \approx 3NI - \frac{N^2}{2kT} \int (|\mathbf{r}_1 - \mathbf{r}_2|) u' e^{-U/kT} d\mathbf{r}_1 \dots d\mathbf{r}_N.$$

Here we have made use of the fact that all $\frac{1}{2}N(N-1)$ terms (the number of interacting pairs) in the double sum are identical with one another. By means of (49.5) and (49.1) we obtain

$$p = \frac{NkT}{V} - \frac{N^2}{6V} \int u'(|\mathbf{r}_1 - \mathbf{r}_2|) d\mathbf{r}_1 d\mathbf{r}_2 \int e^{-U/kT} d\mathbf{r}_3 \dots d\mathbf{r}_N.$$

Making use of (47.7), we finally find

$$p = \frac{NkT}{V} - \frac{N^2}{6V^3} \int \rho_{12}(\mathbf{r}_1, \mathbf{r}_2) (|\mathbf{r}_1 - \mathbf{r}_2|) u' d\mathbf{r}_1 d\mathbf{r}_2. \quad (49.7)$$

Formula (49.7) connects the equation of state with the binary distribution function.

The binary distribution function $\rho_{12}(\mathbf{r}_1, \mathbf{r}_2)$ characterizes the probability of a given mutual disposition of two arbitrarily chosen particles in a system. In isotropic phases (gases and liquids) one more simplification of eq. (49.7) can be made.

Since in isotropic phases the binary function cannot depend on directions but only on the distance between the particles, it can be written in the form

$$\rho_{12}(\mathbf{r}_1, \mathbf{r}_2) = \mu(|\mathbf{r}_1 - \mathbf{r}_2|).$$

Hence

$$p = \frac{NkT}{V} - \frac{N^2}{6V^3} \int \mu(|\mathbf{r}_1 - \mathbf{r}_2|) (|\mathbf{r}_1 - \mathbf{r}_2|) u'(|\mathbf{r}_1 - \mathbf{r}_2|) d\mathbf{r}_1 d\mathbf{r}_2.$$

Introducing a new variable

$$r = |\mathbf{r}_1 - \mathbf{r}_2|$$

one can write

$$\int \mu(|\mathbf{r}_1 - \mathbf{r}_2|) (|\mathbf{r}_1 - \mathbf{r}_2|) u'(|\mathbf{r}_1 - \mathbf{r}_2|) d\mathbf{r}_1 d\mathbf{r}_2 = 4\pi V \int_0^\infty \mu(r) u'(r) r^3 dr,$$

so that finally

$$p = \frac{NkT}{V} - \frac{2\pi N^2}{3V^2} \int_0^\infty \mu(r)u'(r)r^3 dr. \quad (49.8)$$

In an analogous way one can find an expression for the energy of the system:

$$E = kT^2 \frac{\partial \ln Z}{\partial T} = kT^2 \frac{1}{Z_{\text{kin}}} \frac{\partial Z_{\text{kin}}}{\partial T} + \frac{kT^2}{I} \frac{\partial I}{\partial T} = \frac{3NkT}{2} + \frac{kT^2}{I} \frac{\partial I}{\partial T}, \quad (49.9)$$

where $\frac{3}{2}NkT$ is the energy of a system of non-interacting particles.

Substituting the value of I from (47.2) and using (47.7), we calculate $\partial I / \partial T$ to be

$$\frac{\partial I}{\partial T} = \frac{N^2 I}{2V^2 kT^2} \int u(|\mathbf{r}_1 - \mathbf{r}_2|) \rho_{12}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2.$$

Hence

$$E = \frac{3NkT}{2} + \frac{2\pi N^2}{V} \int_0^\infty \mu(r)u(r)r^2 dr. \quad (49.10)$$

Thus the energy as well as the pressure are expressed in terms of the binary distribution function $\mu(r)$.

The binary distribution function $\mu(r)$ can be calculated for gases when the density of particles in the system is small.

That is, in the case of such gases the equation for the binary function (48.4) can be solved by a method of successive approximations.

Indeed, in gases with a density which is not too large the mean distance between particles is large in comparison with their size. The estimations made at the end of §48 show that the coefficient of the threefold distribution function on the right-hand side of (48.4) is proportional to $d^3/(VN^{-1})$ and, consequently, is very small for a sufficiently rarefied gas. This allows one to substitute the approximate value of ρ_{12j} into the integral on the right-hand side of (48.4) without committing any appreciable error.

To obtain an approximate expression for ρ_{12} we expand all distribution functions in a series in powers of the small quantity NV^{-1} — in reality this expansion is carried out in powers of the ratio $[d/(VN^{-1})]^3$ — and we confine ourselves in this expansion to terms of the lowest order, writing

$$\rho_{12} = \rho_{12}^{(0)} + \frac{N}{V} \rho_{12}^{(1)} + \dots, \quad (49.11)$$

$$\rho_{123} = \rho_{123}^{(0)} + \frac{N}{V} \rho_{123}^{(1)} + \dots. \quad (49.12)$$

Substituting these series into the equation for the correlation functions and retaining only low powers of small quantities, one can successively determine the correlation functions, in particular ρ_{12} . Then the pressure, according to (49.8), will appear as a series in powers of NV^{-1} .

$$pV = NkT - \frac{N^2 kT}{2V} \beta - \frac{2}{3} \frac{N^2}{V^2} \beta_2, \quad (49.13)$$

$$\beta = \int (e^{-U/kT} - 1) dV, \quad (49.14)$$

$$\beta_2 = \frac{1}{2} \int (e^{-u(r-r')/kT} - 1)(e^{-u(r)/kT} - 1)(e^{-u(r')/kT} - 1) dV dV'. \quad (49.15)$$

The second term in (49.13) is obviously the same as (46.18), while the third term gives a correction to the pressure to the next smaller order of magnitude (in powers of the density NV^{-1}). In the case of rarefied gases the correlation function method has no special advantages over other methods of calculating corrections for the interaction.

Of greater importance is the application of this method to the construction of a statistical theory of liquids.

Up to now the statistical theory of liquids has been very much in the initial state of its development. The cause of this lies in the very character of thermal motion in liquids.

Thermal motion in liquids differs from that in gases and crystals by the fact that for liquids the energy of the interaction of a molecule with its neighbours cannot be considered to be small (as in gases) or large (as in crystals) in comparison with the energy of thermal motion. For liquids these quantities are of the same order of magnitude.

In a liquid neighbouring molecules oscillate about certain equilibrium positions with a relatively large amplitude. The mutual configuration of molecules is approximately the same as that in an elementary cell of the corresponding crystal.

However, in contrast to crystals, the amplitude of these oscillations is so large that neighbouring molecules draw away from each other relatively easily and leave their equilibrium positions. The mean lifetime τ of a molecule in a given equilibrium position is limited (it is about 10^{-8} sec).

In the course of time intervals which are small in comparison with this time τ , the oscillations of molecules in a liquid have approximately the same character as those in crystals. However, for times $t \gg \tau$ a molecule of a liquid can find itself at any point of the liquid. In this sense its motion is similar to that of a gas molecule.

The character of the jumps and the frequency of the oscillations are determined by the interaction between the molecules in the liquid. This interaction can vary for different liquids over a very wide range. Therefore the basic problem of the contemporary theory of liquids reduces to finding the qualitative characteristics. Such a qualitative characteristic is, in particular, the binary correlation function $\mu(r)$. The binary correlation function characterizes the interaction of the closest neighbours. It can be determined experimentally from the scattering of X-rays.

The correlation function $\mu(r)$ determined in such a way is shown in fig. III.31 by small circles. We see that for a fixed position of a given molecule its closest neighbours are situated with highest probability at distances corresponding to the maxima of the curve $\mu(r)$. It turns out that the positions of these maxima are similar to the corresponding positions in the crystal lattice of the same substance. It is said that short-range order is observed in the disposition of atoms in a liquid and, in this sense, one speaks of the quasi-crystalline structure of water. It is necessary, however, to stress a basic difference between a crystal and a liquid. In liquids the regularity in the disposition of the atoms extends to the closest three or four neighbouring atoms. In crystals the regular disposition of the atoms extends to distances

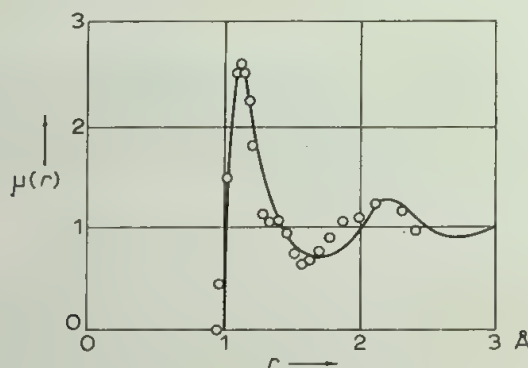


Fig. III.31

which are infinite from the microscopic point of view. In other words, in crystals there is long-range order extending to arbitrarily large distances. As we have already stressed, this difference is associated with the different character of the thermal motion. Hence the analogy between crystals and liquids can be used only within very restricted limits.

By means of certain simplifying assumptions one can obtain from the theory of correlation functions a form of $\mu(r)$ which agrees qualitatively correctly with the experimental data. That is, first, instead of the true energy of interaction a simplified expression is introduced,

$$u(r) = \begin{cases} \infty & \text{if } d \leq 2r \\ 0 & \text{if } d > 2r \end{cases}$$

where d is the diameter of the particles. This energy corresponds to the substitution of hard (impenetrable) spheres of diameter d for the molecules. Second, the so-called superposition approximation is made in the function ρ_{123} . It consists in the substitution

$$\rho_{123} \approx \rho_{13}\rho_{23} . \quad (49.16)$$

The meaning of formula (49.16) is that the interaction of particle 3 with particle 1 is as if particle 2 did not exist at all. In other words, the interaction of particle 3 with particles 1 and 2 is equal to the sum of the pair interactions (31) and (32).

Although the superposition approximation cannot be substantiated theoretically, it is qualitatively clear that it represents an advance over the assumption of the independence of the mutual positions of the particles in space (i.e. over the assumption that $\rho_{123} = \rho_1\rho_2\rho_3$).

In the superposition approximation, eq. (48.4) is closed. It contains only the binary function $\rho_{12} = \mu(r)$.

The solution of the equation for $\mu(r)$ in the approximation mentioned has been carried out numerically. It depends only on one parameter which includes the quantities NV^{-1} , T and d . The solutions obtained are shown in fig. III.31 (by a solid line).

We see that the general forms of the calculated and experimental curves for $\mu(r)$ are very similar. This means that in spite of its schematic character, the model of molecules as hard spheres in the superposition approximation in general correctly describes the character of the interaction of molecules in a liquid.

Crystals

§50. Crystal structure and thermal motion in the one-dimensional crystal model

The contemporary theory of the crystalline state is based on the proposition that the structural units (atoms or molecules) of a crystal are placed at the points of the crystal lattice. Numerous X-ray investigations of crystals and a number of other data have confirmed this proposition and made it possible to measure the distance between atoms in a crystal lattice. In what follows we shall proceed from this proposition as the basis of the theory of the crystalline state.

The distances between atoms in crystals are very small. They are in general of the same order of magnitude as the distances between atoms in molecules, and are sometimes exactly equal to them. For example, the distance between atoms in diamond (1.54×10^{-8} cm) is very close to that between carbon atoms in long chain hydrocarbon compounds (for aliphatic compounds the distance C—C is equal to 1.51×10^{-8} cm).

The distance between molecules in molecular crystals is only two or three times larger than intramolecular distances. Owing to the smallness of these distances the interaction between them is extremely large. In order of magnitude it corresponds to the interaction between atoms in a molecule. From

this point of view an atomic or ionic crystal can be considered as a gigantic molecule containing an enormous number of bound atoms. As in molecules, the energy of interaction between atoms in a crystal is very large in comparison with the energy of thermal motion. Particles in a crystal turn out to be so strongly bound to each other that thermal motion cannot break the bonds.

Thus, from the point of view of the interatomic interaction, crystals represent the opposite limiting case to gases. It is obvious that the only possible form of motion of bound particles in a crystal is the vibrational motion about equilibrium positions. We shall assume that the amplitude of vibrations is very small in comparison with the distances between the atoms. Below we shall discuss in more detail the validity of this assumption.

We shall calculate first of all the mean energy and heat capacity of a crystal whose atoms perform small vibrations about equilibrium positions (lattice points), proceeding from the laws of classical statistics. For this we can make use of the law of equipartition of energy over the degrees of freedom. To each degree of freedom of the motion * there corresponds an energy kT . The number of vibrational degrees of freedom for a crystal containing N atoms is equal to $3N - 6 \approx 3N$ (because N is large). Hence from classical statistics it follows that the mean energy of thermal motion in the crystal is equal to

$$E = 3NkT.$$

The corresponding molar heat capacity is equal to

$$C_V = 3Nk \approx 24.7 \text{ J mol}^{-1}.$$

The heat capacity of crystals turns out to be independent of the temperature and of the concrete properties of crystals and is the same as the well-known Dulong–Petit empirical law for the heat capacity. The Dulong–Petit law yields relatively accurately the heat capacity of many atomic crystals at high temperatures. However, it becomes completely useless on going to low temperatures.

At low temperatures the heat capacity of all crystals decreases with decreasing temperature, as was to be expected from the third law of thermodynamics. Moreover, the heat capacity of certain crystals depends on the temperature even at temperatures which considerably exceed room temper-

* It should be stressed that this holds only for vibrational motion with a small amplitude, when the potential energy is expressed by a quadratic function of the displacement.

ature. As a characteristic example of the total inapplicability of the Dulong-Petit law one can quote diamond. Thus in the case of crystals we again encounter the limited applicability of the law of equipartition, i.e. the limited applicability of classical statistics.

The simplest attempt at applying quantum laws to the treatment of the heat capacity of crystals consists in the following. We consider each atom vibrating at a crystal lattice point as a quantum oscillator which has three degrees of freedom. In a crystal made of atoms of one kind all atoms are completely equivalent and are vibrating with the same frequency ν . If it is assumed that the atoms oscillate independently of each other, then the mean energy of thermal motion E of the entire crystal can be written in the form

$$E = \sum_{n=1}^N \bar{\epsilon}_n,$$

where $\bar{\epsilon}_n$ is the mean energy of the n th oscillator, and the summation is carried out over all oscillators of the crystal. Since each atom is a three-dimensional oscillator, then $\bar{\epsilon}_n = 3\bar{\epsilon}$, where $\bar{\epsilon}$ is the mean energy of the linear quantum oscillator given by formula (43.3).

Thus, the mean energy of a crystal has the form

$$E = \frac{3N h \nu}{2} \coth \frac{h \nu}{2kT},$$

and the heat capacity

$$C_V = \frac{3Nk}{4} \left(\frac{h\nu}{kT} \right)^2 \frac{1}{\sinh^2 (h\nu/2kT)}. \quad (50.1)$$

The form of the heat capacity given by formula (50.1) was discussed in §43 in connection with the vibrational heat capacity of molecules. At high temperatures ($kT \gg h\nu$) the heat capacity tends to the limiting value

$$C_V \approx 3Nk.$$

At low temperatures, in correspondence with the requirement of the third law, C_V tends to zero according to an exponential law:

$$C_V \approx \frac{3Nk}{4} \left(\frac{h\nu}{kT} \right)^2 e^{-h\nu/kT}. \quad (50.2)$$

The simplest quantum law yields qualitatively the correct behaviour of the heat capacity of a crystal with temperature. However, a more detailed comparison of formula (50.2) with experimental data for the heat capacity shows that (50.2) does not give all the features of the behaviour of the heat capacity. The heat capacity of crystals decreases with temperature not according to the exponential law (50.2) but according to a power law of the form $C_V \sim T^3$. The disagreement of formula (50.2) with experiment is associated with the error in the assumption of the independence of the oscillations of atoms in a crystal, which was taken as the basis of the derivation of the formula. In reality, atoms in a crystal are so strongly bound with each other that the idea that the individual motion of one atom is independent of the motion of other atoms is out of the question. The vibrational motion of atoms in a crystal has a collective character, and all the atoms in the crystal simultaneously take part in it.

In order to get a clearer idea of the character of the thermal motion of atoms in a crystal, we shall make use of a fictitious crystal model in the form of a chain of atoms which are distributed along a line at equal distances from each other. Such a chain can be considered as a one-dimensional crystal. Although in nature there are no one-dimensional crystals, the consideration of the thermal motion in a one-dimensional crystal will allow us to elucidate the character of the motion in a real three-dimensional crystal.

We number the atoms in the chain in such a way that the number n may run over values from $n = 1$ up to $n = N$ (there are all together N atoms in the chain). Assume that a certain atom (ion or molecule) having, say, a number n goes out of an equilibrium position and is displaced a distance ξ_n to the right or to the left. Then it will be subject to forces from the neighbouring atoms: a repulsive force on the part of the neighbour it approached, and an attractive force on the part of the other neighbour. Since the forces of intermolecular interaction rapidly decrease with increasing distance, we can take into account only the interaction of the given atom with its two closest neighbours — the atoms with numbers $n - 1$ and $n + 1$. Even the next atoms, with numbers $n + 2$ and $n - 2$, will interact only very weakly with the atom considered, so that this interaction can be disregarded.

The force acting on the n th atom due to each of its two neighbours can be written in the form

$$F_n = - \frac{\partial u(\xi_n)}{\partial \xi_n} ,$$

where $u(\xi_n)$ is the potential energy of the n th atom at the point ξ_n . For

small displacements the potential energy $u(\xi_n)$ can be expanded in a series in powers of a small quantity ξ_n and one can restrict oneself to the first terms of the expansion, as is always done in the theory of small oscillations:

$$u(\xi_n) = u(0) + \left(\frac{\partial u}{\partial \xi_n} \right)_0 \xi_n + \left(\frac{\partial^2 u}{\partial \xi_n^2} \right)_0 \cdot \frac{\xi_n^2}{2} + \dots$$

Since at the point $\xi = 0$ the potential energy has a minimum, then at that point $\partial u / \partial \xi_n = 0$ and $\partial^2 u / \partial \xi_n^2 = \kappa > 0$. The force acting on the particle is

$$F = -\kappa \xi_n.$$

Here it has been assumed that the neighbouring atoms having numbers $n-1$ and $n+1$ remained at rest at their points of the crystal lattice. In reality, of course, this is not so. A displacement of the n th atom will lead to a displacement of the $(n-1)$ th and $(n+1)$ th atoms; the $(n+1)$ th atom will be displaced under the action of a repulsive force to the right a distance ξ_{n+1} , while the $(n-1)$ th atom under the action of an attractive force will follow the n th atom and will be displaced a distance ξ_{n-1} . Hence the distance between the n th atom and its neighbours will be changed respectively by $\xi_{n+1} - \xi_n$ and $\xi_{n-1} - \xi_n$. The n th atom will be acted upon by a force

$$F_n = \kappa(\xi_{n+1} - \xi_n) + \kappa(\xi_{n-1} - \xi_n) = \kappa(\xi_{n+1} + \xi_{n-1} - 2\xi_n). \quad (50.3)$$

The displacement of the $(n+1)$ th atom will lead to a displacement of the $(n+2)$ th, while the displacement of the $(n-1)$ th atom will lead to a displacement of the $(n-2)$ th. These atoms will in their turn act on the next neighbours, and as a result the entire chain of atoms will be set in motion. To investigate this motion it is sufficient to find the motion of an arbitrarily chosen n th atom. The equations of its motion have the form

$$m\ddot{\xi}_n = \kappa(\xi_{n+1} + \xi_{n-1} - 2\xi_n). \quad (50.4)$$

We shall assume that the first and N th atoms of the chain are fixed, so that their displacements are

$$\xi_1 = \xi_N = 0. \quad (50.5)$$

Of course, we have no special grounds for this assumption. However, from the general propositions of statistical physics it follows that the motion of a

system consisting of a very large number of particles cannot depend on the character of the initial and boundary conditions. Hence the condition which we have imposed upon two atoms of a chain containing N (for $N \gg 1$) atoms cannot have an essential effect on the motion of the entire chain. If, instead of the condition (50.5), we introduced the conditions

$$\frac{\partial u(\xi_1)}{\partial \xi_1} = \frac{\partial u(\xi_N)}{\partial \xi_N} = 0$$

which would express the fact that the first and last atoms of the chain remain free and that the force acting on them reduces to zero, the final result for $N \gg 1$ would not be changed at all.

As we have already said, the departure of any atom from an equilibrium position gives rise to a perturbation which propagates along the chain. This perturbation moves from atom to atom until it reaches the last atom which is fixed at the end of the chain. Here the perturbation will not vanish but will be reflected and will propagate in the opposite direction to the other end of the chain. At the other end it will again be reflected, and so on. In the chain of atoms there will arise waves travelling in opposite directions. The superposition of these waves will lead to the formation of standing waves similar to those arising in an elastic string with fixed ends. We shall seek the particular solution of (50.4) in the form

$$\xi_{nf} = A e^{i\omega t} e^{ifan}, \quad (50.6)$$

where A is the complex amplitude of the wave, and a is the distance between neighbouring equilibrium positions. The substitution of ξ_{nf} into (50.4) gives

$$m\omega^2 = \kappa(2 - e^{+ifa} - e^{-ifa}),$$

whence for ω we find

$$\omega = (\kappa/m)^{\frac{1}{2}} [2(1 - \cos fa)]^{\frac{1}{2}} = 2(\kappa/m)^{\frac{1}{2}} \sin \frac{1}{2}fa. \quad (50.7)$$

This formula, which connects the frequency ω and the wave number f , is the law of dispersion of waves in the chain. The wave numbers must be determined from the boundary conditions (50.5) which correspond to standing waves. The latter can be written in the form of a linear combination of the expression (50.6):

$$\xi_n = A \sin fx \sin(\omega t + \alpha) = A \sin fan \sin(\omega t + \alpha), \quad (50.8)$$

where a is the distance between equilibrium positions, A is the amplitude of the wave, and α is the phase. The conditions (50.5) at the ends of the chain will be satisfied if one sets $\sin afN = 0$, or

$$f = \frac{\pi k}{aN}, \quad (50.9)$$

where the number k runs over a series of integers: $k = 1, 2, 3, \dots, N$. Thus, the displacement ξ_n can be written in the form of a superposition of waves of the form

$$\xi_{nk} = A_k \sin \frac{\pi kn}{N} \sin(\omega_k t + \alpha). \quad (50.10)$$

It follows from formula (50.10) that the appearance of N standing waves with different frequencies ω_k ($k = 1, 2, \dots, N$) is possible in a chain of N atoms. To these N frequencies there correspond N wave numbers (50.8) or N different wavelengths:

$$\lambda_k = \frac{2\pi}{f_k} = \frac{2Na}{k}. \quad (50.11)$$

The longest of the standing wave (for $k=1$) has a wavelength $2Na$, i.e. one half-wave can be fitted into the entire chain. To the values $k = 1, 2, \dots$ there correspond ever shorter waves such that an integer number of half-waves is fitted into the length of the chain. It is remarkable that all atoms of the chain vibrate with the same frequency (i.e. ω_k depends only on k but not on the ordinal number of the atom).

The velocity of propagation of the waves is equal to

$$v_k = \frac{\omega_k}{f_k} = \frac{\omega_k \lambda_k}{2\pi} = v_k \lambda_k = 2 \left(\frac{\kappa}{m} \right)^{\frac{1}{2}} \frac{\sin \frac{1}{2} f_k a}{f_k}, \quad (50.12)$$

i.e. turns out to be different for different waves. For small wave numbers f_k , i.e. for long waves, $\sin \frac{1}{2} f_k a$ can be expanded in a series and we can write $\sin \frac{1}{2} f_k a \approx \frac{1}{2} f_k a$. In this case

$$\omega_k = \left(\frac{\kappa}{m} \right)^{\frac{1}{2}} f_k a, \quad (50.13)$$

$$v = v_0 = a \left(\frac{\kappa}{m} \right)^{\frac{1}{2}} = \text{const.} \quad (50.14)$$

In the case of very long waves, for which the inequality $f_k a \ll 1$ or $\lambda \gg a$ is fulfilled, a very large number of atoms are oscillating almost in phase; a change of the phase takes place in a half wavelength which covers a large number of atoms. Hence the atomic structure of the lattice has no effect on its properties. The lattice behaves with respect to long waves as a continuous elastic medium. Standing waves in the lattice are equivalent to those in an elastic medium. The velocity of propagation of waves in the lattice, given by (50.14), is the same as that of elastic waves (the velocity of sound) in a continuous medium.

To the shortest possible waves, $\lambda_N = 2Na/N = 2a$, there corresponds the highest frequency:

$$\omega_N = 2 \left(\frac{\kappa}{m} \right)^{\frac{1}{2}} \quad (50.15)$$

and the velocity of propagation

$$v = v_{\min} = \frac{2a}{\pi} \left(\frac{\kappa}{m} \right)^{\frac{1}{2}}. \quad (50.16)$$

Comparison of formulae (50.14) and (50.16) shows that long waves propagate with a velocity which is somewhat larger than that of short waves, i.e. the phenomenon of dispersion takes place. In the intermediate frequency range the velocity of propagation of waves, determined by formula (50.12), depends on the wave number f_k or the frequency ω_k .

In addition to the boundary condition (50.5) use is often made of the so-called conditions of periodicity. If an entire infinite chain of atoms is divided into segments of length L each containing N atoms ($L=aN$) then the motion of all the segments must be the same. This means that the following conditions must be fulfilled:

$$\xi_n = \xi_{n+N}$$

or, upon substitution of ξ_n from (50.6),

$$e^{ifaN} = 1.$$

Hence

$$f = \pm \frac{\pi k}{aN}, \quad (50.9')$$

where k takes on even integer values $k = 2, 4, \dots, N$. Here f has positive as well as negative values in the interval $-\pi/a \leq f \leq \pi/a$. The interval between the neighbouring values f_k and f_{k+1} doubles. The total number of waves remains the same.

The function $\omega(f)$ is shown in fig. III.32.

The displacement of an arbitrary n th atom in the chain is given in the form of a superposition of displacements of the form (50.10), i.e.

$$\xi_n = \sum_k \xi_{nk} = \sum_k A_k \sin(\omega_k t + \alpha) \sin f_k a n, \quad (50.17)$$

where the summation is carried out over all possible values of the wave vector f_k . Instead of an arbitrary amplitude A_k we introduce the amplitude

$$C_k = \left(\frac{N-1}{2} \right)^{\frac{1}{2}} A_k$$

and let

$$C_k \sin(\omega_k t + \alpha) = q_k. \quad (50.18)$$

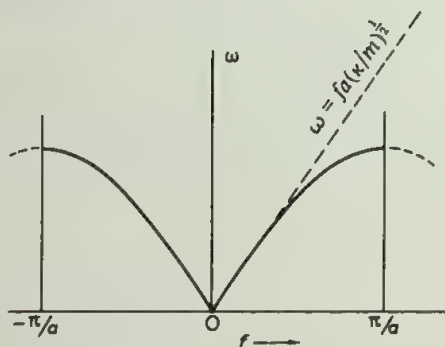


Fig. III.32

Then we have

$$\xi_n = \left(\frac{2}{N-1} \right)^{\frac{1}{2}} \sum_k q_k \sin f_k a n. \quad (50.19)$$

If the values of the amplitudes q_k are given, then it follows from formula (50.19) that the displacement of an arbitrary atom in the chain will be completely determined. Hence the amplitudes q_k can be considered as the generalized coordinates of the system.

We now find the energy of the entire chain expressed in terms of the generalized coordinates q_k . The kinetic energy of the chain is obviously equal to

$$T = \frac{1}{2} m \sum_n \dot{\xi}_n^2,$$

where the summation is carried out over all atoms of the chain. We have

$$\begin{aligned} \dot{\xi}_n^2 &= \frac{2}{N-1} \left(\sum_k \dot{q}_k \sin f_k a n \right)^2 = \\ &= \frac{2}{N-1} \sum_k \sum_{k'} \dot{q}_k \dot{q}_{k'} \sin f_k a n \sin f_{k'} a n, \end{aligned}$$

hence

$$T = \frac{m}{2} \sum_n \sum_k \sum_{k'} \frac{2}{(N-1)} \dot{q}_k \dot{q}_{k'} \sin f_k a n \sin f_{k'} a n.$$

Changing the order of summation, we obtain

$$T = \frac{m}{2} \sum_k \sum_{k'} \frac{2}{N-1} \dot{q}_k \dot{q}_{k'} \sum_n \sin f_k a n \sin f_{k'} a n.$$

By virtue of the orthogonality property of sine functions, the sum $\sum_n \sin f_k a n \sin f_{k'} a n$ is equal to zero, provided $f_k \neq f_{k'}$, i.e. $k \neq k'$. If $f_k = f_{k'}$, then $k = k'$ and

$$\sum_{n=1}^{N-1} \sin f_k a n \sin f_{k'} a n = \frac{N-1}{2}.$$

Hence

$$T = \frac{1}{2} m \sum_k \dot{q}_k^2. \quad (50.20)$$

The kinetic energy of the chain is expressed by the quadratic form of the derivatives of the coordinates q_k .

We now find the potential energy of the entire chain. From formula (50.3) it follows that the potential energy of the entire crystal is equal to

$$U = \sum u_n = \frac{1}{2} \kappa \sum (\xi_{n+1} - \xi_n)^2, \quad (50.21)$$

where u_n is the potential energy of the n th atom.

We can convince ourselves of this by differentiating eq. (50.21). Indeed, if the force F_n acting on the n th atom is

$$F_n = - \frac{\partial U}{\partial \xi_n} = - \frac{\kappa}{2} \frac{\partial}{\partial \xi_n} [(\xi_2 - \xi_1)^2 + \dots + (\xi_n - \xi_{n-1})^2 + (\xi_{n+1} - \xi_n)^2 + \dots + (\xi_N - \xi_{N-1})^2] = \kappa(\xi_{n-1} + \xi_{n+1} - 2\xi_n),$$

then from formula (50.19) we find

$$\begin{aligned} \xi_{n+1} - \xi_n &= \left(\frac{2}{N-1} \right)^{\frac{1}{2}} \sum_k q_k \{ \sin [f_k a(n+1)] - \sin f_k a n \} = \\ &= \left(\frac{2}{N-1} \right)^{\frac{1}{2}} \sum_k q_k 2 \cos(f_k a n + \frac{1}{2}) \sin \frac{1}{2} f_k a. \end{aligned}$$

Proceeding in the same way as in calculating the kinetic energy, we have

$$\begin{aligned} U &= \frac{\kappa}{2} \frac{2}{N-1} 4 \sum_n \sum_k \sum_{k'} q_k q_{k'} \cos(f_k a n + \frac{1}{2}) \times \\ &\times \cos(f_{k'} a n + \frac{1}{2}) \sin \frac{1}{2} f_k a \sin \frac{1}{2} f_{k'} a = \frac{\kappa}{2} \frac{2}{N-1} 4 \times \\ &\times \sum_k \sum_{k'} q_k q_{k'} \sin \frac{1}{2} f_k a \sin \frac{1}{2} f_{k'} a \times \sum_n \cos(f_k a n + \frac{1}{2}) \cos(f_{k'} a n + \frac{1}{2}). \end{aligned}$$

But for cosines the orthogonality condition holds:

$$\sum \cos(f_k a n + \frac{1}{2}) \cos(f_{k'} a n + \frac{1}{2}) = \begin{cases} \frac{N-1}{2} & \text{for } k = k' , \\ 0 & \text{for } k \neq k' , \end{cases}$$

hence

$$U = \frac{1}{2} \kappa \sum_k q_k^2 (2 \sin \frac{1}{2} f_k a)^2 .$$

Taking into account formula (50.7), we find finally

$$U = \frac{1}{2} m \sum_k \omega_k^2 q_k^2 . \quad (50.22)$$

Thus, the total energy of the crystal is equal to

$$E = \frac{1}{2} m \sum_k (\dot{q}_k^2 + \omega_k^2 q_k^2) . \quad (50.23)$$

Formula (50.23) has an important physical meaning: the total energy of the crystal is expressed in a quadratic form containing only the squares of the quantities \dot{q}_k and q_k (but not their products of the form $\dot{q}_k q_k$). Hence the quantities q_k are the normal coordinates of the vibrating crystal. Each term in (50.23) has the form

$$\epsilon = \frac{1}{2} m (\dot{q}_k^2 + \omega_k^2 q_k^2) = \frac{1}{2} m (\dot{q}_k^2 + 4\pi^2 \nu_k^2 q_k^2) , \quad (50.24)$$

i.e. represents the energy of a linear harmonic oscillator with a mass equal to that of an atom oscillating with a frequency ν_k . The energy E is equal to the sum of the energies of such oscillators which have different frequencies ν_k . The energy of a crystal of N atoms which perform bound oscillations turns out to be equal to the energy of N independent harmonic oscillators with a set of frequencies ν_k determined by formula (50.7). In this sense a system of N atoms which perform bound oscillations is equivalent to a set of N independent oscillators with frequencies ν_k . Instead of finding the mean energy of a complex system of N bound atoms, we can seek the mean energy of a

much simpler equivalent system of N independent oscillators. It should be stressed that linear oscillators with an energy given by formula (50.24) have nothing in common with real atoms (with the exception of the same mass). Each oscillator represents one of the normal oscillations of the crystal as a whole. All atoms oscillating with one and the same frequency ν_k take part in the normal oscillation of the crystal.

The transition which we have made from displacements ξ_n to normal coordinates q_k represents a transformation which is typical for wave processes, and is not connected with the properties of the linear chain. The possibility of the transition to normal coordinates, in which the energy has the quadratic form (50.23), represents a general algebraic theorem.

Real crystals made of atoms (or molecules) of the same mass m are very seldom encountered. Usually the crystal lattice contains particles with different masses and different chemical natures. The latter leads to a change in the law of interaction between neighbouring atoms, so that the quantity κ has different values at different points of the crystal.

Without taking into account the above effect we shall consider only the effect of the difference in the masses of the atoms on the character of the motion in the case of the one-dimensional-chain model.

Let atoms with masses m_1 and m_2 having one and the same physical nature (i.e. with the same values of the quasi-elastic constant) be placed alternately at equal distances in a chain. The equations of motion of the chain will now be written in the form

$$m_1 \ddot{\xi}_n = \kappa(2\xi_n - \eta_{n+1} - \eta_{n-1}), \quad (50.25)$$

$$m_2 \ddot{\eta}_n = -\kappa(2\eta_n - \xi_{n+1} - \xi_{n-1}), \quad (50.26)$$

where ξ_n and η_n are the displacements of the atoms with masses m_1 and m_2 respectively.

Analogously to (50.6), we can seek the solution of eqs. (50.25) and (50.26) which satisfies the boundary conditions (50.5) in the form

$$\xi_{nf} = A e^{ifan} e^{i\omega t},$$

$$\eta_{nf} = B e^{ifan} e^{i\omega t}.$$

The substitution of these expressions into (50.25) and (50.26) leads to the system of algebraic equations

$$(-\omega^2 m_1 + 2\kappa)A = 2\kappa \cos(fa)B,$$

$$(-\omega^2 m_2 + 2\kappa)B = 2\kappa \cos(fa)A.$$

Eliminating the amplitudes, one can obtain the following expression for ω^2 :

$$\omega^2 = \kappa \left(\frac{1}{m_1} + \frac{1}{m_2} \right) \pm \kappa \left[\left(\frac{1}{m_1} + \frac{1}{m_2} \right)^2 - \frac{4 \sin^2 fa}{m_1 m_2} \right]^{\frac{1}{2}} \quad (50.27)$$

If ω is written as a function of f , then depending on the sign in front of the root one obtains the two branches shown in fig. III.33. For small f , expanding the square root in (50.28) in a series in powers of f , we obtain

$$\omega_- \approx \left[\frac{2\kappa}{m_1 + m_2} \right]^{\frac{1}{2}} (fa), \quad (50.28)$$

$$\omega_+ \approx \left[2\kappa \left(\frac{1}{m_1} + \frac{1}{m_2} \right) \right]^{\frac{1}{2}}. \quad (50.29)$$

The branch ω_- , called the acoustic branch, corresponds to $\omega \rightarrow 0$ as $f \rightarrow 0$. Its behaviour does not differ significantly from that of the dispersion curve (50.13) for a chain of identical atoms.

The second branch of frequencies, which is absent in a chain of identical atoms, shows a completely different behaviour (the upper curve in fig. III.33). As $f \rightarrow 0$, ω_+ tends to the constant limit (50.29). This branch of frequencies is

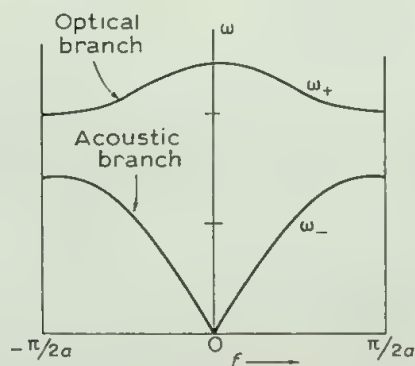


Fig. III.33

called the optical branch. For given f the frequency of the waves of the optical branch is much higher than that of the acoustic branch.

For small f it is easy to determine the ratio of the wave amplitudes of the optical branch:

$$\frac{A}{B} \approx -\frac{m_2}{m_1}. \quad (50.30)$$

The minus sign in the above formula shows that particles with masses m_1 and m_2 which are involved in optical waves are moving towards each other.

In acoustic waves neighbouring particles with different masses are moving in the same direction.

We see that the difference in the masses of the atoms change fundamentally the character of thermal motion in the crystal.

Consider the case when at the points of the crystal lattice there are charged ions instead of neutral atoms. The optical branch corresponds to motion of the ions towards each other. Hence optical waves give rise to variations in the polarization of the crystal, and it is optical waves that have an effect on electromagnetic processes in crystals. Their name is associated with this fact.

§51. Long waves in a three-dimensional crystal

Thermal motion in a three-dimensional crystal has, in general, the same character as in the one-dimensional model. A displacement of an arbitrary atom from the equilibrium position in the lattice is transmitted to its closest neighbours in three dimensions. Their displacements will, in their turn, give rise to displacements of other atoms, and an elastic wave propagating in three dimensions will arise in the crystal. As a result of the reflection of the elastic waves from the faces of the crystal a system of standing waves will be established in the crystal. In the same approximation as for the one-dimensional model (disregarding the third powers of displacements) the energy of a crystal can be written in normal coordinates in the form

$$E = \sum \left(\frac{m\dot{q}_k^2}{2} + \frac{4\pi^2 m \nu_k^2 q_k^2}{2} \right), \quad (51.1)$$

where the summation is carried out over all possible wave numbers. The bound oscillations of atoms in the three-dimensional crystal are equivalent to a set of $3N$ independent linear oscillators with the natural frequencies ν_k .

The determination of the natural frequencies for the three-dimensional crystal presents very great mathematical difficulties. Hence in order to find the thermodynamic functions of a crystal it is necessary to make further simplifying assumptions in addition to the proportionality of the forces to the first power of the displacements. First, in considering thermal waves in a crystal we shall confine ourselves to the case of long waves ($\lambda \gg a$). As we have seen in the preceding paragraph, in the case of long waves very large groups of atoms are oscillating in phase and one can disregard the discrete atomic structure of the crystal. Similarly in the case of long waves in a three-dimensional crystal one can digress from the discrete structure and consider the crystal as a continuous elastic medium. Second, we shall disregard the anisotropy of the crystal and shall consider it as an isotropic elastic medium. In the isotropic elastic medium the thermal perturbations of the crystal form a system of standing waves. In a three-dimensional elastic medium the wave number f is equal to

$$f = (f_1^2 + f_2^2 + f_3^2)^{\frac{1}{2}}, \quad (51.2)$$

where f_1 , f_2 and f_3 are quantities which characterize respectively the propagation of the wave in three mutually perpendicular directions.

In order that a condition of the type (50.5), i.e. the condition of reflection of elastic waves from the faces of the crystal may be fulfilled, the wave numbers must satisfy the following conditions:

$$f_1 = \frac{\pi k_1}{aN}; \quad f_2 = \frac{\pi k_2}{aN}; \quad f_3 = \frac{\pi k_3}{aN}, \quad (51.3)$$

where k_1 , k_2 and k_3 are integers (1, 2, ..., N). The frequency of long waves is connected with the wave numbers by a relation which represents a direct generalization of formula (50.13).

In contrast to the one-dimensional model, in a three-dimensional isotropic elastic medium the propagation of three elastic waves is possible: one longitudinal wave (in which displacements take place in the direction of propagation of the wave) and two transverse waves (in which displacements are perpendicular to the direction of propagation). The velocities of propagation c_l and c_t of the longitudinal and transverse waves are different. Hence, instead of (50.12), in the three-dimensional case one must write

$$2\pi\nu_l = c_l f, \quad 2\pi\nu_t = c_t f,$$

where ν_l and ν_t are the frequencies of the longitudinal and transverse elastic waves, and c_l and c_t are their velocities.

In what follows we shall need to know the number of elastic waves whose frequency lies in the interval between ν_l and $\nu_l + d\nu_l$ and ν_t and $\nu_t + d\nu_t$ respectively. The calculation of this quantity in no way differs from the calculations in §38 of Part I.

Indeed, in §38 of Part I the number of travelling waves in a cavity of volume V has been calculated. The number of standing waves in a crystal can be found in a completely analogous way. The only difference from the calculation carried out in §38 of Part I lies in the fact that in a crystal wave vectors are determined by formula (51.3) which differs from (38.11) of Part I by the absence of the factor 2. On the other hand, here the integers k_1, k_2, k_3 take on only positive values, whereas in §38 of Part I they take on negative as well as positive values.

As a result, for the number of longitudinal waves with a frequency between ν_l and $\nu_l + d\nu_l$ one obtains the formula

$$g(\nu_l) d\nu_l = \frac{4\pi V}{c_l^3} \nu_l^2 d\nu_l, \quad (51.4)$$

which is identical with formula (38.22) of Part I *.

For the number of transverse waves with a frequency between ν_t and $\nu_t + d\nu_t$ we find analogously

$$g(\nu_t) d\nu_t = 2 \frac{4\pi V}{c_t^3} \nu_t^2 d\nu_t,$$

where the factor 2 appears because in an elastic medium there are two transverse waves with one and the same frequency ν_t . The total number of elastic waves whose frequency lies between ν and $\nu + d\nu$ is obviously equal to

* It should be noted for what follows that the displacement represented in formula (50.18) in the form of a standing wave can be written in the form of a superposition of two travelling waves. Reproducing the calculations of §38 of Part I one can write ξ in the form

$$\xi = N^{-\frac{1}{2}} \sum_f \sum_{j=1}^3 e_{ff}(q_{ff}) e^{ifr + q_{ff}^*} e^{-ifr}. \quad (51.5)$$

$$g(\nu)d\nu = 4\pi V \left(\frac{1}{c_l^3} + \frac{2}{c_t^3} \right) \nu^2 d\nu. \quad (51.6)$$

Since we have compared each wave with a frequency ν_k to an oscillator oscillating with the same frequency, the quantity $g(\nu)d\nu$ which we have calculated represents the number of oscillators whose frequency lies between ν and $\nu + d\nu$. If the crystal were of infinite size and contained an infinitely large number of atoms, the number of possible frequencies or oscillators would also be infinite. In reality, however, it is equal to $3N$. Therefore we can write

$$3N = \sum g(\nu_k), \quad (51.7)$$

where the summation is carried out over all possible frequencies.

We have established the form of the function $g(\nu)$ in the range of long waves or small frequencies, in which it can be assumed to be a continuous function of the argument ν given by formula (51.6). However, in the range of high frequencies the form of the spectral function is unknown and depends on the concrete structure of a given crystal.

Debye proposed a method of calculating the thermodynamic functions of crystals which is, in essence, based on a certain interpolation.

Namely, the spectral function $g(\nu)$ is assumed to have the form (51.6) over the entire frequency range, and in the entire frequency range the summation is replaced by integration. However, the integration is carried out up to a certain limiting frequency ν_{\max} which is expressed in terms of the number of particles in the crystal by means of the condition (51.7). Thus, the spectral function $g(\nu)$ is assumed to have the form

$$g(\nu) = \begin{cases} 4\pi V \left(\frac{1}{c_l^3} + \frac{2}{c_t^3} \right) \nu^2, & \nu \leq \nu_{\max}, \\ 0, & \nu > \nu_{\max}. \end{cases} \quad (51.8)$$

This gives

$$3N = \int_0^{\nu_{\max}} 4\pi V \left(\frac{1}{c_l^3} + \frac{2}{c_t^3} \right) \nu^2 d\nu = 4\pi V \left(\frac{1}{c_l^3} + \frac{2}{c_t^3} \right) \frac{\nu_{\max}^3}{3}, \quad (51.9)$$

whence

$$\nu_{\max} = \left(\frac{9}{4\pi} \frac{N}{V} \frac{c_l^3 c_t^3}{2c_l^3 + c_t^3} \right)^{\frac{1}{3}}. \quad (51.10)$$

The largest, or limiting, frequency ν_{\max} turns out to depend only on quantities which can be measured experimentally — the velocities of sound c_l and c_t — and to be proportional to the density of the crystal $(N/V)^{\frac{1}{3}}$.

The introduction of such a “cut-off” spectral function (51.8) leads to expressions for thermodynamic functions which in the limiting cases of low and high temperatures go over into accurate expressions, and in the range of intermediate temperatures have the character of interpolation formulae.

By means of the expression for the limiting frequency ν_{\max} determined by formula (51.10) the spectral function $\nu(g)$ can be rewritten in the more compact form

$$g(\nu)d\nu = \frac{9N}{\nu_{\max}^3} \nu^2 d\nu. \quad (51.11)$$

§ 52. The partition function of a crystal

We have established in the preceding paragraphs that the thermal motion in a crystal containing N atoms is described by a set of $3N$ independent oscillators whose frequencies lie between zero and ν_{\max} . In order to find the partition function of the entire crystal it is necessary to find the partition function of a system consisting of $3N$ independent oscillators. Since they are independent, we can, obviously, write this function in the form of the product of the partition functions of all the oscillators, i.e.

$$Z = \prod_{k=1}^{3N} z_k, \quad (52.1)$$

where Z is the partition function of the crystal, and z_k is the partition function of the individual k th oscillator. It should be stressed that the oscillator is not an individual atom but characterizes a definite oscillation of the crystal as a whole. Hence in (52.1) one should not carry out the division by $(3N)!$, as should be done in the case of $3N$ identical independent particles. We have

calculated the partition function z_k of a quantum oscillator in §43. Taking the logarithm of (52.1) and substituting (43.2) into it, we find

$$\ln Z = \sum_{k=1}^{3N} \ln z_k = \sum_{k=1}^{3N} \ln \frac{\exp(-h\nu_k/2kT)}{[1 - \exp(-h\nu_k/kT)]}. \quad (52.2)$$

To calculate the sum (52.2) it is necessary to know all the possible frequencies ν_k of the crystal.

However, as we have already mentioned, this problem is still unsolved. Therefore we shall confine ourselves to the Debye approximation and shall replace the summation by the integration over "cut-off" spectrum (51.11). This gives

$$\begin{aligned} \ln Z &= \int_0^{\nu_{\max}} \ln \frac{\exp(-h\nu/2kT)}{[1 - \exp(-h\nu/kT)]} g(\nu) d\nu = \\ &= -\frac{9N}{\nu_{\max}^3} \int_0^{\nu_{\max}} \frac{h\nu^3}{2kT} d\nu - \frac{9N}{\nu_{\max}^3} \int_0^{\nu_{\max}} \ln(1 - e^{-h\nu/kT}) \nu^2 d\nu. \end{aligned} \quad (52.3)$$

To calculate the integrals in formula (52.3) we introduce a new variable

$$x = h\nu/kT, \quad (52.4)$$

and the characteristic temperature θ_c of the crystal:

$$\theta_c = h\nu_{\max}/k,$$

which is analogous to the characteristic temperature introduced in §43. We obtain

$$\ln Z = -\frac{9N\theta_c}{8T} - 9N \left(\frac{T}{\theta_c} \right)^3 \int_0^{\theta_c/T} x^2 \ln(1 - e^{-x}) dx. \quad (52.5)$$

The calculation of the above integral can be carried out only for low and high temperatures. We shall understand low temperatures to be temperatures which are considerably lower than the characteristic temperature θ_c of the crystal. For $T \ll \theta_c$ the limit of the integral can be replaced by infinity, since the integrand is very small for any large values of the argument x . This gives

$$\int_0^{\theta_c/T} x^2 \ln(1-e^{-x}) dx \approx \int_0^{\infty} x^2 \ln(1-e^{-x}) dx. \quad (52.6)$$

This integral (52.6) is calculated in Appendix IV. Substituting its value into (52.5), we have

$$\ln Z = -\frac{9N\theta_c}{8T} + \frac{\pi^4 N}{5} \left(\frac{T}{\theta_c}\right)^3. \quad (52.7)$$

The substitution of infinity for the limit of the integral (52.6) has an important physical meaning. It shows that at $T \ll \theta_c$ only vibrations with small frequencies ν are excited in the crystal (i.e. the small values of x are essential). For large frequencies (large x) the integrand reduces to zero, and the corresponding frequencies give no contribution to the value of Z . This justifies the approximation made in the preceding paragraph — the replacement of the discrete crystal by a continuous elastic medium in which only oscillations with small values of ν are excited.

For high temperatures $T \gg \theta_c$ the limit of the integral will be a small quantity. Hence x in the integrand is small, and the integrand can be expanded in a series in powers of x . In this case we have

$$\ln(1-e^{-x}) \approx \ln x,$$

so that

$$\int_0^{\theta_c/T} x^2 \ln(1-e^{-x}) dx \approx \int_0^{\theta_c/T} x^2 \ln x dx = \frac{1}{3} \left(\frac{\theta_c}{T}\right)^3 \ln \frac{\theta_c}{T} - \frac{1}{9} \left(\frac{\theta_c}{T}\right)^3. \quad (52.8)$$

Substituting formula (52.8) into (52.5), we find

$$\ln Z = -3N \ln \frac{\theta_c}{T} + N - \frac{9}{8} N \left(\frac{\theta_c}{T}\right). \quad (52.9)$$

By means of the expressions (52.7) and (52.9) one can find the thermodynamic functions of a crystal at high and low temperatures.

§ 53. The thermodynamic functions of a crystal

We shall first of all calculate the energy and heat capacity of a crystal at low temperatures. At low temperatures ($T \ll \theta_c$) the energy is equal to

$$E = kT^2 \frac{\partial \ln Z}{\partial T} = \frac{9}{8} Nk\theta_c + \frac{3\pi^4}{5} \frac{NkT^4}{\theta_c^3}. \quad (53.1)$$

The first term in formula (53.1) represents the energy of the crystal as $T \rightarrow 0$, i.e. the zero-point energy. The second term shows that the energy of the crystal increases rapidly (as T^4) with increasing temperature.

The heat capacity of the crystal at low temperatures is given by the formula

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = \frac{12\pi^4 Nk}{5} \left(\frac{T}{\theta_c} \right)^3. \quad (53.2)$$

It turns out to be proportional to the cube of the absolute temperature.

A characteristic feature of the expressions (53.1) and (53.2) is the fact that they involve a material constant of the crystal: its characteristic temperature θ_c . Hence at low temperatures different crystals possess different heat capacities (which are smaller the higher θ_c).

At high temperatures ($T \gg \theta_c$) the energy and heat capacity are equal respectively to

$$E = kT^2 \frac{\partial \ln Z}{\partial T} = 3NkT + \frac{9}{8} Nk\theta_c \approx 3NkT, \quad (53.3)$$

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = 3Nk. \quad (53.4)$$

As was to be expected, the values of E and C_V agree with those obtained from the law of equipartition of energy over degrees of freedom. They do not depend on the material constants of the crystal and are universal quantities. The independence of the energy and heat capacity on the material constants is due to the fact that at sufficiently high temperatures the energy of the crystal turns out to be independent of the vibrational frequencies of the crystal. This fact allows one to understand why formulae (53.3) and (53.4) turn out to be correct in spite of the fact that they are derived on the basis of an undoubtedly incorrect assumption. Indeed, in deriving them it was

assumed that the basic role is played by small frequencies (long waves), for which the discrete crystal can be considered as a continuous elastic medium. For $T \gg \theta_c$ in addition to low frequencies there must also be excited in the crystal high frequencies, which will give a considerable contribution to the partition function. Hence the approximate law of frequency distribution (51.8) will no longer be applicable. However, in the classical approximation, which is valid for sufficiently high temperatures, the value of Z and, consequently, also that of the energy of the crystal depend neither on the frequencies themselves nor on the character of their distribution.

In the intermediate range of temperatures $T \approx \theta_c$ the energy and heat capacity are expressed by more complex formulae, which are obtained by numerical integration of formula (52.5). Since the width of the transitional range is not large, they are of no special interest.

It should be noted that, although we have calculated the heat capacity at constant volume, the expression obtained agrees to a high degree of accuracy with the heat capacity at constant pressure, because in a solid body the two heat capacities are practically the same.

We now find the entropy S of the crystal. For low temperatures ($T \ll \theta_c$) from (53.1) and (52.7) we get

$$S = \frac{E}{T} + k \ln Z = \frac{4\pi^4}{5} Nk \left(\frac{T}{\theta_c} \right)^3. \quad (53.5)$$

For high temperatures ($T \gg \theta_c$) from (53.3) and (52.9) we obtain

$$S = \frac{E}{T} + k \ln Z = 3Nk \ln \frac{T}{\theta_c} + 4Nk. \quad (53.6)$$

Formula (53.6) agrees with the purely thermodynamic expression for the entropy and differs from the latter only in the fact that it does not contain an indefinite constant entropy. Formula (53.5) shows that as $T \rightarrow 0$ the entropy of an atomic crystal tends to zero as the cube of the temperature, which is in complete agreement with the requirements of the third law of thermodynamics, just as is the decrease of the heat capacity as $T \rightarrow 0$.

Finally, we find the free energy of the crystal.

From the conditions (52.7) and (52.9) we obtain

$$F = -kT \ln Z \approx -N \frac{\pi^4 kT}{5} \left(\frac{T}{\theta_c} \right)^3 + \frac{9Nk\theta_c}{8} \quad \text{for} \quad T \ll \theta_c, \quad (53.7)$$

$$F \approx -3NkT \ln \frac{T}{\theta_c} - NkT \quad \text{for} \quad T \gg \theta_c. \quad (53.8)$$

Proceeding from the free energy one can obtain the equation of state of the crystal, which has the following form for $T \gg \theta_c$:

$$p = - \left(\frac{\partial F}{\partial V} \right)_T = - \frac{3NkT}{\theta_c} \frac{\partial \theta_c}{\partial V}.$$

Since p turns out to be expressed in terms of the quantity $\partial \theta_c / \partial V$ and the latter cannot be calculated theoretically or found from sufficiently simple experiments, this equation of state is of no great practical importance.

It is interesting to find the law of distribution of energy levels in a solid body or, more precisely, the spacing between the energy levels. If $\Omega(\epsilon)$ is the number of levels per unit interval of energy, then the spacing between neighbouring energy levels is obviously equal to

$$D(\epsilon) = \Omega^{-1}(\epsilon).$$

By means of formulae (24.10) and (53.5) one can write

$$D(\epsilon) = e^{-\sigma} = e^{-S/k} = \exp \left[-\frac{4}{3} \pi^4 N (T/\theta_c)^3 \right].$$

The spacing between levels decreases with decreasing temperature as well as with decreasing number of particles. For temperatures which are very close to the absolute zero it turns out, in accordance with what was said in §35, to be equal to kT even for a macroscopic body.

§54. Comparison of theory with experiment

For the practical use of the expressions obtained and the comparison of calculated values with experimental data it is necessary to know the characteristic temperature θ_c of the crystal.

From the definition and formula (51.10) we find

$$\theta_c = \frac{h\nu_{\max}}{k} = \frac{h}{k} \left(\frac{9N}{4\pi V} \right)^{\frac{1}{3}} \left(\frac{c_1^3 c_t^3}{2c_1^3 + c_t^3} \right)^{\frac{1}{3}}. \quad (54.1)$$

Formula (54.1) contains in addition to numerical values and universal constants, two quantities which are determined by the properties of the crystal: the density N/V and the velocity of sound in the crystal. These quantities have been measured for many crystals. Since the velocity of sound, according to (50.14), is inversely proportional to the square root of the mass of the atoms, it is particularly large for very hard crystals made of light atoms. For such crystals the characteristic temperature is particularly high.

Table 8 gives the characteristic temperatures of certain crystal. It follows from this table that for crystals of the type lead and common salt room temperature (300 K) and higher temperatures (up to 1000 K) are relatively, though not very, high. Hence for such crystals the departures from classical laws and, in particular, from the classical value of the heat capacity are not very large in this region. Without much error, it can be assumed that their heat capacity is equal to 24 J/mol. However, at temperatures which are lower than 100 K the departure from this value becomes very appreciable for all these crystals.

Table 8

Crystal	θ_c (K)	Crystal	θ_c (K)
Pb	88	Cu	315
I	106	Be	1000
benzene	150	Al	398
Na	172	Fe	453
Ag	215	diamond	1860
NaCl	281		

The situation is different for crystals with a high characteristic temperature, particularly in the case of diamond. For the latter room temperature appears to be low, and the applicability of classical laws is out of the question. The heat capacity of diamond already follows the T^3 law at room temperature.

Measurements of the variation of the heat capacity with temperature, which have been carried out for a very large number of crystals, show that the theory is in very good agreement with experiment. At temperatures $T \ll \theta_c$ the heat capacities indeed follow the T^3 law and tend to zero as $T \rightarrow 0$. At $T \approx \theta_c$ a gradual transition to the classical (constant) value of the heat capacity takes place. The complete form of the curve of the heat capacity of some crystals is shown in fig. III.34. The small circles on the curve show the measured values of the heat capacity of different crystals.

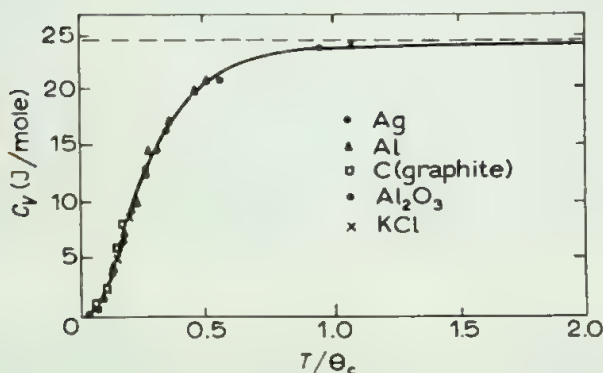


Fig. III.34

They are all presented on the universal scale T/θ_c . The agreement of theory with experiment proves to be very good.

It should, however, be kept in mind that the rough theory developed holds only for crystals made of particles whose internal structure can be disregarded. This means that one can disregard the effect of the temperature on the state of the particles. In most cases this condition is satisfied by crystals made of atoms. In atoms the spacing between the ground state and the first excited state is as a rule large in comparison with kT , and thermal motion cannot have an effect on the state of atoms. Hence their internal energy does not depend on the temperature and they give no contribution to the heat capacity. However, for certain atoms the lowest electronic levels lie very close to each other. Thus, for example, for the ions of gadolinium which enter into crystalline gadolinium sulphate the lowest energy level consists of eight sublevels which are spaced at distances corresponding in the temperature scale to the characteristic temperature 1.6 K. At very low temperatures, $T \approx 7$ K, in accordance with the results of §40, there appears an additional heat capacity superposed on the heat capacity of the crystal lattice. Since at such low temperatures the heat capacity of the crystal lattice is already very small, the increase of the heat capacity is very sharp. At $T = 1.6$ K the additional heat capacity exceeds that of the crystal lattice by a factor of almost 500. As the temperature decreases further the heat capacity of the system falls to zero.

In the case of crystals made of complex molecules the internal structure of the particles cannot, as a rule, be disregarded. In the first approximation one

can disregard the effect of the vibration of a molecule in the crystal lattice on its internal thermal motion. The molecule as a whole is vibrating in the crystal lattice, while vibrations of individual atomic groups take place inside it. In certain cases the internal motion of the molecule represents a free rotation. Thus, for example, H_2 molecules must be considered as rotating freely in the hydrogen crystal. The rotating H_2 molecules take part simultaneously in the thermal vibrations of the lattice. Disregarding the interaction of the internal motion in molecules with the motion of the molecule as a whole, the total energy E of the crystal can be written in the form

$$E = E_{\text{lat}} + E_{\text{intern}} ,$$

where E_{lat} is the mean energy of the vibrations of the crystal lattice, and E_{intern} is the mean internal energy of the molecule, which we have calculated in ch. 5. Correspondingly, the heat capacity of the crystal can be written in the form

$$C_V = C_{V_{\text{lat}}} + C_{V_{\text{intern}}} .$$

The contribution of the internal motion to the heat capacity can be very substantial in certain cases. Thus, for example, the heat capacity of intramolecular vibrations in benzene amounts to about 20% of the heat capacity of the lattice at $T \approx 150$ K and reaches about 80% of the latter at $T \approx 270$ K. Hence in calculating the heat capacity of complex crystals it is necessary to take into account the contribution of the internal motion, particularly at high temperatures. Good agreement of the theory with experiment justifies the simplifications made.

However, in a number of cases disagreement, though not considerable, is observed between theory and experiment. At high temperatures vibrational amplitudes become large, and it is no longer justified to disregard the squares of displacements in the expression for quasi-elastic forces *. The vibrational motion loses its simple harmonic character. The error which arises as a result of the assumption of the isotropy of the elastic medium is relatively negligible. The error arising because the discrete character of the crystal is disregarded turns out to be more considerable. The effect of the distribution of high frequencies, which we have disregarded in determining the number of the

* In this connection it should be noted that, as can be shown by calculation, a crystal for which the forces were exactly proportional to the displacements would have a coefficient of thermal expansion equal to zero.

normal vibrations of the crystal, shows up in the fact that for hard crystals the characteristic temperature θ_c turns out not to be a constant but a function of the temperature of the crystal. For example, in the case of lithium its value changes from 330 K at a temperature of the crystal equal to 20 K to 410 K at a temperature of 120 K. A similar phenomenon occurs in the case of diamond. All the errors mentioned are so negligible that they have only been noticed relatively recently because of the increased accuracy of measurements.

It should also be noted that the existence of anomalies at low temperatures, which are similar to those existing for gadolinium, can lead to seeming contradictions with the third law of thermodynamics. Usually the value of the heat capacity as $T \rightarrow 0$ is found by an extrapolation from values which are measured at higher temperatures. If this extrapolation is carried out from a certain temperature which is higher than that at which the increase of the heat capacity occurs, then a considerable error arises. The value of $\int_0^T T^{-1} C_V dT$ obtained from the extrapolation can differ appreciably from the experimental value of the entropy at a high temperature found from other data. Hence in order to find the true values of thermodynamic functions it is advisable to carry out the measurements of the heat capacity at temperatures as low as possible.

The Theory of Fluctuations

§55. Small fluctuations in macroscopic systems

In the preceding discussions we have more than once pointed to a difference between the statistical and purely thermodynamic ideas on the nature of thermal processes. From the laws of statistical physics the existence of fluctuations necessarily follows. A system undergoing a fluctuation can pass spontaneously over from a more probable state into one of the less probable states. In this case the trend of the process is the reverse of that for which an increase in entropy occurs.

The probability of fluctuations in a closed system can be calculated by means of the Boltzmann formula. Simple estimations carried out by means of this formula, as well as the general considerations discussed in §36, show that the probability of any appreciable fluctuations in a system which contains a large number of particles is extremely low. The phenomenon of fluctuations can in practice be observed in two cases: (1) when the dimensions of the system are sufficiently small; in this case fluctuations will occur often and their scale will be relatively large; (2) when the dimensions of the system are not small, but rather small fluctuations occur. Such small fluctuations may occur often, but the departure of the system from an equilibrium state will be relatively small. In this chapter we shall consider both of these cases of fluctuations.

In order to estimate correctly the role which the investigations of fluctuations played in the development of molecular-statistical concepts, it should be kept in mind that the existence of fluctuations was predicted theoretically at the time when the second law of thermodynamics seemed to many to be one of the dogmas of physics. The representatives of the so-called school of energeticians denied the existence of material atoms and molecules. Statistical physics, in which the laws of classical mechanics were unified with statistical laws, seemed to be internally inconsistent, and was accepted with distrust by many physicists. Hence the discovery of numerous examples of fluctuation processes was a brilliant confirmation of the laws of statistical physics and was one of the most important events in the final establishment of molecular theory. In the studies by Einstein and Smoluchowski it was shown that a number of physical processes, which had been known for a long time, are due to fluctuation phenomena, and a quantitative theory of these processes was developed which proved to be in excellent agreement with experimental facts. The significance of these discoveries can best be expressed in the words of Smoluchowski himself *:

“At present we do not regard the dogmas of physics with the same esteem as before. Great changes concerning the problem of the significance of kinetic atomistics and thermodynamics have taken place. They are associated with the fact that we have only recently managed to interpret, on the basis of kinetic theory, facts which were known a long time ago; for example, Brownian motion discovered as far back as 1827, the phenomenon of critical opalescence discovered more than 20 years ago, the well-known blue coloration of the sky, and so on. The new thing which we encounter in these interpretations, and which is in contradiction with every-day established notions, lies in the fact that they were the first to take seriously into account the Maxwell velocity distribution law. As a result they were the first to consider heat as a process of motion, whereas before this the concept of the nature of heat was usually considered as a kind of poetic simile.”

We shall begin the consideration of fluctuation processes with the second case, i.e. the case of systems whose dimensions are large.

We shall below expound the general theory of small fluctuations occurring in an arbitrary macroscopic system. We take a closed system which is in a state of statistical equilibrium having an entropy S_0 . We now assume that the state of the system changes in such a way that it passes over into a non-equilibrium state in which its entropy is equal to S . We assume that the change in the state of the system can be characterized by a change in a certain internal

* M.Smoluchowski, Phys. Z. 13 (1912) 1059.

parameter ξ whose value depends on the state of the entire system. In an equilibrium state the parameter ξ has a value $\xi = \xi_0$, whereas in a non-equilibrium state its value differs from ξ_0 .

As an example of the parameter ξ one can take the density ρ of a gas confined in a closed, thermally insulated container. In an equilibrium state the density is constant over the entire volume of the container, i.e. $\xi_0 = \rho_0 = \text{const}$. As a result of a fluctuation the system can spontaneously go over into a non-equilibrium state with a variable density $\xi = \rho(x)$. Other examples will be discussed later.

The entropy of the system will be a function of the parameter ξ , so that one can write that $S = S(\xi)$. In an equilibrium state $S_0 = S(\xi_0)$. The probability that the closed system considered will get into a state characterized by a value of the parameter ξ which lies in the interval between ξ and $\xi + d\xi$ can be found by means of the Boltzmann formula:

$$dw = \text{const} \cdot \exp\left(\frac{S(\xi) - S(\xi_0)}{k}\right) d\xi = \text{const} \cdot \exp\left(\frac{\Delta S}{k}\right) d\xi, \quad (55.1)$$

where the constant is determined by the normalization condition *. The value of the change in the entropy is obviously negative.

The applications of formula (55.1) to actual cases of fluctuations will be considered in the following section. Formula (55.1) is applicable to fluctuations in a system with a constant energy.

However, one very often has to consider fluctuations occurring not in a closed, but a quasi-closed system which constitutes a small part of a closed system. Such a quasi-closed system can be considered as a subsystem placed in a reservoir with a constant temperature T_0 . We shall assume that fluctuations occur only in the subsystem, while the reservoir is always in an equilibrium state. The state of the subsystem will be characterized by the value of a certain external parameter λ . In going over from an equilibrium state to a non-equilibrium state the parameter λ changes from λ_0 to λ . As λ changes, the values of the thermodynamic quantities characterizing the subsystem also change. We shall assume that the variations of the macroscopic parameter λ are sufficiently slow, so that at every instant an equilibrium statistical distribution exists in the subsystem. Then we can consider that the thermodynamic quantities in the subsystem are interrelated by the usual

* Strictly speaking, the constant in (55.1) also depends on the parameter ξ . It can, however, be shown that for a system containing a sufficiently large number of particles the dependence on ξ of the factor which stands in front of the exponential is insignificant in comparison with the dependence on the exponential.

equilibrium relations. The process of the transition from an equilibrium state into a non-equilibrium state in a subsystem placed in a reservoir can be considered as a transition performed under the action of a certain external source of work. As the parameter λ changes by an amount $\Delta\lambda = \lambda - \lambda_0$ the source does work $\Delta W(\lambda)$ on the subsystem.

We now write an expression for the probability that the subsystem will go over into a state with a value of λ between λ and $\lambda + \Delta\lambda$ while the reservoir remains in an equilibrium state. Since the reservoir and subsystem together constitute a closed system, formula (55.1) is applicable to them. However, in it the change in the entropy must be written in the form

$$\Delta S = \Delta S_0 + \Delta S',$$

where $\Delta S'$ is the change in the entropy of the subsystem. Then the probability that the subsystem will go over into a state with λ lying in the interval $\lambda, \lambda + d\lambda$ under the action of the external source of work is given by the formula

$$dw = \text{const} \cdot \exp\left(\frac{\Delta S_0 + \Delta S'}{k}\right) d\lambda. \quad (55.2)$$

But, by virtue of our assumption of the slowness of the variation of the macroscopic parameters, one can write for $\Delta S'$ the usual equilibrium expression:

$$\Delta S' = \frac{\Delta E' + p_0 \Delta V' - \Delta W}{T_0}, \quad (55.3)$$

where T_0 and p_0 are the equilibrium temperature and equilibrium pressure of the system (which are equal to the corresponding quantities of the reservoir), and E' and V' are the energy and volume of the subsystem. (In the last formula it is seen clearly that ΔW represents the work done by an external source and not by the reservoir. The work done by the reservoir is equal to $-p_0 \Delta V'$.) Further,

$$\Delta S_0 = \frac{\Delta E_0 + p_0 \Delta V_0}{T_0}.$$

But by virtue of the fact that the system is closed (reservoir+subsystem) the total volume of the system remains constant, so that

$$\Delta V_0 = -\Delta V'.$$

The energy conservation law gives

$$\Delta E' + \Delta E_0 = 0,$$

hence

$$\Delta S_0 = -\Delta S' - \frac{\Delta W(\lambda)}{T_0}. \quad (55.4)$$

Substituting (55.4) into (55.2), we find

$$dw = \text{const} \cdot \exp \left(-\frac{\Delta W(\lambda)}{kT_0} \right) d\lambda. \quad (55.5)$$

Thus, in the most general case, it can be said that the work which must be done on a macroscopic system in order to change the parameter λ , characterizing the state of the system, by an amount $\Delta\lambda$ is a measure of the probability of small fluctuations in the system. This does not mean, however, that a system can undergo a fluctuation only when real work from without is done on it. This is particularly clearly seen from the example of a closed system on which no work is done. The work ΔW is only a quantitative characteristic of a fluctuation. The work ΔW can be written as the change in the potential energy as the system is displaced in a certain imaginary (and sometimes also real) field of force. Denoting the potential of this field by $u(\lambda)$, we have

$$\Delta W = u(\lambda) - u(\lambda_0) = u(\lambda),$$

if $u(\lambda_0)$ is chosen as the zero-point potential energy. Then formula (55.5) can be written in the form

$$dw = \text{const} \cdot \exp \left(-\frac{u(\lambda)}{kT_0} \right) d\lambda = w(\lambda) d\lambda. \quad (55.6)$$

We arrive thus at a formula which is an analogue of the Boltzmann formula. In what follows we shall see that this analogy has a completely clear meaning.

To calculate the probability of a fluctuation according to formulae (55.5) or (55.6) it is necessary in each individual case to find the work done or the change in the potential energy which takes place in the fluctuation process.

By virtue of the smallness of the fluctuations the expression for $u(\lambda)$ can be expanded in a power series of the small parameter $\lambda - \lambda_0$, and one can confine oneself to the first terms of the expansion:

$$u(\lambda) = u'(\lambda_0)(\lambda - \lambda_0) + \frac{1}{2}u''(\lambda_0)(\lambda - \lambda_0)^2 + \dots,$$

where the primes denote the derivatives with respect to λ . In an equilibrium state the potential energy of the field must have a minimum, so that

$$u'(\lambda_0) = 0 \quad \text{and} \quad u''(\lambda_0) > 0.$$

Hence the probability distribution (55.6) can be written in the form

$$dw = \text{const} \cdot \exp\left(-\frac{u''(\lambda_0)(\lambda - \lambda_0)^2}{2kT_0}\right) d\lambda. \quad (55.7)$$

The probability distribution (55.7) is called the Gaussian distribution. The value of the constant $u''(\lambda_0)$ depends on the nature of that real or fictitious field of force in which the system is "displaced" from the position λ_0 to the position λ . By means of the probability distribution of small fluctuations (55.7) one can find the mean value of the fluctuation of the parameter λ :

$$\Delta^2 = \overline{(\lambda - \lambda_0)^2} = \text{const} \int (\lambda - \lambda_0)^2 \exp\left(-\frac{u''(\lambda_0)(\lambda - \lambda_0)^2}{2kT_0}\right) d\lambda.$$

The constant in (55.7) is determined by the normalization condition

$$\text{const} = \left[\int \exp\left(-\frac{u''(\lambda_0)(\lambda - \lambda_0)^2}{2kT_0}\right) d\lambda \right]^{-1}.$$

Thus,

$$\Delta^2 = \frac{\int (\lambda - \lambda_0)^2 \exp\left(-\frac{u''(\lambda_0)(\lambda - \lambda_0)^2}{2kT_0}\right) d\lambda}{\int \exp\left(-\frac{u''(\lambda_0)(\lambda - \lambda_0)^2}{2kT_0}\right) d\lambda}. \quad (55.8)$$

The fluctuations of the parameter λ occur in both directions from its value in an equilibrium state. Since the integrand in the integrals in the numerator and denominator of the expression (55.8) rapidly decreases with increasing

absolute value of the difference $\lambda - \lambda_0$, the integration can be carried out in the range from $-\infty$ to $+\infty$, as we have done in normalizing the Maxwell distribution. Thus, finally,

$$\Delta^2 = \frac{\int_{-\infty}^{\infty} (\lambda - \lambda_0)^2 \exp \left(-\frac{u''(\lambda_0)(\lambda - \lambda_0)^2}{2kT_0} \right) d\lambda}{\int_{-\infty}^{\infty} \exp \left(-\frac{u''(\lambda_0)(\lambda - \lambda_0)^2}{2kT_0} \right) d\lambda} = \frac{kT_0}{u''(\lambda_0)}. \quad (55.9)$$

By means of formula (55.9) the probability distribution (55.7) can be written in the form

$$dw = (2\pi\Delta^2)^{-\frac{1}{2}} \exp \left(-\frac{(\lambda - \lambda_0)^2}{2\Delta^2} \right) d\lambda.$$

The probability of a given fluctuation decreases sharply with its increasing value, as well as with decreasing Δ^2 . The latter quantity is proportional to the absolute temperature. Hence it can be stated that the intensity of fluctuations decreases with decreasing temperature *.

In the following sections the general relations obtained will be applied to concrete cases of small fluctuations in macroscopic systems.

§56. Brownian motion

As the first case where fluctuations turn out to be easily observable we shall consider the so-called Brownian motion. Brownian motion is the continuous chaotic motion of small particles suspended in a liquid or a gas which may be observed under a microscope.

A complete quantitative theory of Brownian motion, which not only explained its nature but also allowed one to predict a number of its characteristic features, was developed in the studies of Einstein and Smoluchowski (1905–1906). The investigations of Brownian motion played a major role in the triumph of the molecular-kinetic theory because Brownian motion was the first physical process in which the existence of molecules was detected in a direct and obvious way. The significance and importance of the theory of Brownian motion are not confined to the historical aspect. On the contrary, it

* For an exception to this rule see B.G. Levich, *Vvedenie v statisticheskuyu fiziku (Introduction to statistical physics)* (Gostekhizdat, Moscow, 1954) § 63.

is only relatively recently that a number of cases of Brownian motion have become of particular interest in association with the creation of new, very accurate measuring devices (see § 58).

Passing on to the analysis of the theory of Brownian motion, we shall consider a macroscopic particle suspended in a volume of liquid or gas, and shall seek the forces acting on it due to the molecules of the medium. The molecules of the medium are in continuous thermal motion. Hence the molecules of the liquid or gas in which the particle is suspended will continually collide with the particle and transfer momentum to it in each collision. In other words, the molecules of the medium will exert a pressure on the surface of the particle. The collisions of the molecules with the surface of the particle are completely random, from all directions. If the surface of the particle is sufficiently large, so that a large number of molecules impinge on it in a very short time interval, then it can be assumed that the momenta which are transferred to the particle from all directions are, on the average, balanced. The situation is different in the case of very small particles (with a size of the order of 10^{-4} cm). Such particles still contain an enormous number of molecules and are macroscopic bodies. Nevertheless, the surface of such particles is so small that in a short time a relatively small number of molecules impinge on it. The resultant of the forces exerted by the molecules of the medium on the surface of the particle turns out to be different from zero. As a result the particle will be given a random motion whose direction and velocity will vary at a very high frequency (of the order of 10^{12} times per second). The character of the displacements is shown in fig. III.35, in which the position

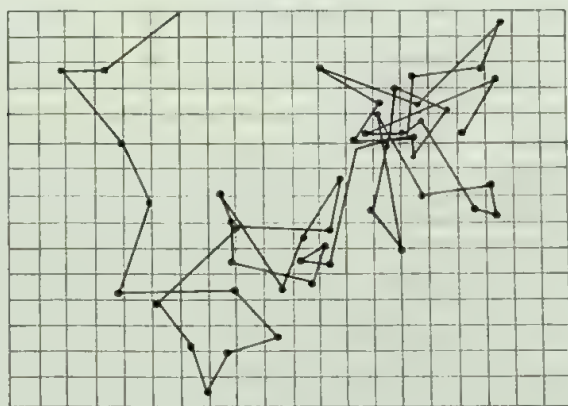


Fig. III.35

of the particle every 30 seconds is drawn. The side of each square corresponds to a distance of 3×10^{-4} cm.

The number of molecules striking the particle and the momentum which they transfer to the particle will undergo large fluctuations. Thus, the Brownian motion is due to the fluctuations of the pressure exerted by the molecules of the medium on the particle suspended in it. Although its motion is not directly a molecular motion, it serves as a kind of indicator of molecular motion. As we have more than once stressed, the phenomena of fluctuations contradict the propositions of pure thermodynamics. This can be illustrated particularly clearly by the example of Brownian motion.

The very fact of the persistence of Brownian motion and the impossibility of abolishing it points to the continuous violation of the requirements of the second law of thermodynamics. Indeed, if a particle suspended in a medium obtained a single momentum from any external source, then its motion would rapidly be slowed down as a result of the energy loss due to viscous friction. Hence the impossibility of abolishing Brownian motion is indicative of the existence of processes which are the reverse of the processes of viscous friction and which are accompanied by a decrease in the entropy. To maintain its motion the particle continuously draws energy from the medium surrounding it, which directly contradicts the second law of thermodynamics.

In order to construct a quantitative theory of Brownian motion we can make use of the general relations introduced in the preceding paragraph.

Let a very small but macroscopic particle with mass μ be suspended in a certain medium, a liquid or a gas. We assume that the position of this particle is characterized by a certain parameter (a generalized coordinate) λ . Such a parameter can be, for example, the distance of the particle from a certain plane of the container which is chosen as the origin (other examples will be given below). The particle will be acted upon on the part of the medium by a force which varies rapidly and randomly in time and is due to the fluctuations of the thermal motion of the molecules of the medium. Under the action of this fluctuation force, which we shall call for brevity Brownian force, the particle will undergo very small displacements, so that the value of its parameter λ will vary continually by very small amounts $\Delta\lambda$.

Instead of observing the motion of one particle in time one can, following Einstein, consider a great number of identical particles undergoing Brownian displacements, and find the number of particles passing through a certain imaginary surface in the medium.

We denote by $c(\lambda)$ the number of particles per unit volume which are at a distance between λ and $\lambda + d\lambda$ from the surface $\lambda = 0$.

Let $\Delta = [(\Delta\lambda)^2]^{1/2}$ denote the mean square displacement of particles in a

certain short time τ . Then on the average $[\frac{1}{2}c(\lambda - \frac{1}{2}\Delta)]\Delta$ particles moving from left to right will pass through 1 cm^2 of an imaginary surface in the solution during the time τ . Analogously, $[\frac{1}{2}c(\lambda + \frac{1}{2}\Delta)]\Delta$ particles moving in the opposite direction will pass through this surface in the same time. As a result the following number of particles will pass through 1 cm^2 of the imaginary surface:

$$N = j\tau = [\frac{1}{2}c(\lambda - \frac{1}{2}\Delta) - \frac{1}{2}c(\lambda + \frac{1}{2}\Delta)]\Delta, \quad (56.1)$$

where j is the particle flux.

Assuming Δ to be small and $c(\lambda)$ to be a slowly varying function of the coordinate λ , we can write

$$N \approx -\frac{\Delta^2}{2} \frac{\partial c}{\partial \lambda}$$

or

$$j = -\frac{\Delta^2}{2\tau} \frac{\partial c}{\partial \lambda}. \quad (56.2)$$

The flux of particles is proportional to the gradient of their concentration and is directed in the direction of the decrease of the concentration. The factor of proportionality $\Delta^2/2\tau$ is called the diffusion coefficient D :

$$D = \frac{\Delta^2}{2\tau}. \quad (56.3)$$

Thus, the mean square displacement of the particle is equal to

$$\Delta^2 = \overline{(\Delta\lambda)^2} = 2D\tau. \quad (56.4)$$

The mean part traversed by the particle turns out to be proportional to the square root of the observation time τ .

The diffusion coefficient D can be expressed in terms of the temperature and physicochemical constants of the medium. That is, we assume that the flux of particles is due not only to the gradient of concentration but also to an external force \mathbf{f} acting on every one of the particles.

Under the action of a force \mathbf{f} a small particle in a viscous medium moves with a velocity \mathbf{u} which for steady motion is equal to

$$\mathbf{u} = b\mathbf{f} = \frac{\mathbf{f}}{C\eta a}, \quad (56.5)$$

where $b = (C\eta a)^{-1}$ is a quantity called the mobility of the particle, a is its radius, η is the viscosity of the medium, and C is a numerical coefficient equal to 6π for spherical particles. Formula (56.5) is called the Stokes formula. A simple calculation shows that the time required to establish steady motion is very short for small particles.

The total flux of particles can be written in the form

$$j = -D \frac{\partial c}{\partial \lambda} + uc \quad (56.6)$$

or

$$j = -D \frac{\partial c}{\partial \lambda} + bfc = -D \frac{\partial c}{\partial \lambda} - bc \frac{\partial U}{\partial \lambda}, \quad (56.7)$$

where U is the potential energy corresponding to the force \mathbf{f} .

We now assume that the flux of particles caused by the external field has the same value as that produced by the gradient of concentration but the opposite direction. Then the total flux of particles j reduces to zero. In this case the particle concentration distribution is determined by the condition

$$-D \frac{\partial c}{\partial \lambda} - bc \frac{\partial U}{\partial \lambda} = 0$$

or

$$c = c_0 e^{-bU/D}. \quad (56.8)$$

On the other hand, we know that particles which do not interact with each other follow, in an external field, the Boltzmann distribution, and

$$c = c_0 e^{-U/kT}. \quad (56.9)$$

Comparing these last expressions, we find

$$D = bkT. \quad (56.10)$$

Thus, the diffusion coefficient of particles is connected with their mobility by the universal formula (56.10), which for spherical particles assumes the form

$$D = \frac{kT}{6\pi\eta a}. \quad (56.11)$$

Substituting the value of D from (56.11) into (56.4), we find

$$\Delta = [(\Delta\lambda)^2]^{1/2} = \left(\frac{kT}{3\pi\eta a}\right)^{1/2} \tau^{1/2}. \quad (56.12)$$

Thus, the mean path traversed by particles increases with the temperature of the medium.

All the quantities contained in formula (56.12) are known or can be measured. It should be noted that the value of the Boltzmann constant k was at one time determined from formula (56.12). The degree of agreement of these formulae with experiment can be estimated from the fact that in 1910–1915 the value of Avogadro's number $N = R/k$ found from measurements of Brownian motion ($N = 6.44 \times 10^{23}$) was considered to be one of the most accurate values of this quantity.

Experiments on Brownian motion demonstrated in a direct and obvious way one more important inference of statistical mechanics. That is, the statement on the reversibility of molecular processes. The experiments consisted of observing the number of Brownian particles in a sharply limited (for example, by the corresponding illumination) field of view under a microscope. Owing to the Brownian motion, particles enter and leave the field of observation from the non-illuminated part of the solution. Let us assume that at a certain instant of time the concentration of Brownian particles in the field of observation was higher than in the remaining solution. According to the second law of thermodynamics equalization of concentration must then take place by means of the diffusion of particles from the illuminated volume to the non-illuminated one. After the final equalization of concentrations in the system a total equilibrium must be established which should not subsequently be violated. From the point of view of statistical physics the phenomenon should proceed in a completely different way. The number of particles in a sufficiently small volume should increase and decrease equally frequently, so that the notions of diffusion and equalization of concentrations would lose any meaning. After the lapse of a recovery time τ^* the number of particles which was initially equal, say, to n , should come back

to the same value. The length of the recovery time was calculated by Smoluchowski. As we have mentioned already, it increases sharply with the size of the system, in the case given with the value of the number n . The results of the observations are given in tables 9 and 10.

Table 9
Observed frequency of variation of the number of particles $n \rightarrow m$ in the viewing field

n	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$
0	210	126	35	7	0	1	—
1	134	281	117	29	1	1	—
2	27	138	108	63	16	3	—
3	10	20	76	38	24	6	0
4	2	2	14	22	13	11	3
5	—	0	2	10	10	1	3

Table 10
Mean recovery time

Observation time	$\tau^*(\text{obs.})$	$\tau^*(\text{cal.})$
0	6.1	5.5
1	3.1	3.2
2	4.1	4.0
3	7.8	8.1
4	18.6	20.9

In the first table the frequency of variation of the number of particles in the field of observation is given. The number of particles in this field in the first observation is denoted by n , while m denotes the number of particles in a subsequent observation. The frequency of a transition $n \rightarrow m$ means the number of cases in which n was replaced by m . For example, the number 27 in the third row of the second column of table 9 means that in 27 cases the number of particles which in the first observation was equal to two in the second observation reduced to zero. The mean number of particles which should have been found in the field of view amounted to $\bar{n} = 1.43$. The measurements were carried out in time intervals $\Delta t = 1.39$ sec.

An analysis of the numbers of table 9 indicates immediately the correct-

ness of the statistical point of view, and serves as a direct illustration of the reasoning of § 25. Indeed, according to the propositions of thermodynamics, we should expect a continuous decrease of the number of particles ($m < n$) in the case where the initial value $n > \bar{n}$, and an increase of the number of particles in the reverse case. Nothing of the sort is found in table 9. On the contrary, for $n > \bar{n}$, in subsequent observations an even larger number of particles is very often observed. Thus, for $n = 3$ a smaller number of particles ($m=0,1,2$) is observed in 106 cases in the second observation, and a larger or the same number of particles ($m=3,4,5$) is observed in 68 cases.

From table 9 it is seen that for a small number of particles the numbers which stand on the two sides of the major diagonal are practically equal to each other. For example, the frequency of the transition from $n = 3$ to $m = 0$ amounts to 10. The frequency of the transition from $n = 0$ to $m = 3$ is equal to 7. The frequency of the transition from $n = 2$ to $m = 4$ is equal to 16, while that from $n = 4$ to $m = 2$ is equal to 14, and so on. This means that the process of Brownian motion has a strictly reversible character. Fluctuations occur so frequently that no systematic trend of them with time is observed. If, however, the number of particles n is considerable, so that the scale of fluctuations is large, then in correspondence with the reasoning of § 25 a resolution of the fluctuations can be expected. In this case the most probable trend of the process is the same as that predicted by thermodynamics: particles will most often go out (diffuse) of the observation zone, and the number of particles in it should in most cases decrease.

From table 9 it is seen that for $n = 5$ (such an n already sufficiently substantially exceeds \bar{n}) the number of particles decreases in 22 cases and only in 4 cases does it increase or remain constant. If the number of particles n were very large and very much exceeded the mean number \bar{n} , then it would decrease in an overwhelming majority of cases, and the process would become irreversible.

The data of table 10 are no less convincing. This table gives the calculated and observed times of recovery of the number of particles in the observation field (in units of $\Delta t = 1.39$ sec) for a suspension with a mean number of particles $\bar{n} = 1.55$. From the table it is seen that after the lapse of time intervals τ^* , which are in good agreement with those calculated theoretically, the number of particles initially found in the field of view again recurs. The recovery time increases sharply with the value of the departure of n from \bar{n} , so that large fluctuations are very seldom repeated (see also table 9). All these facts are convincingly indicative of the validity of the molecular-statistical point of view.

§57. Fluctuations of thermodynamic quantities in a homogeneous system

We shall now consider the fluctuations of thermodynamic quantities with reference to a system in a reservoir.

A quantitative measure of the probability of a fluctuation is the work which must be done on the subsystem in order to bring it from the initial equilibrium state into the final fluctuation state.

Because of the smallness of the fluctuations the transitions can be assumed to be reversible.

The work done in a reversible transition for a system in a medium is expressed by the general thermodynamic formula (28.7):

$$\Delta W = \Delta E - T_0 \Delta S + p_0 \Delta V, \quad (57.1)$$

where ΔE , ΔS and ΔV are the changes in the corresponding quantities for the transition from the initial to the final state. A concrete expression for the work ΔW can be obtained for particular cases of the process.

We shall confine ourselves to the calculation of this work for fluctuations of the volume at a constant temperature and fluctuations of the temperature at a constant volume.

We shall first of all consider the fluctuations of the volume at a constant temperature ($T = T_0 = \text{const}$).

The work done in an isothermal change in the volume at a constant temperature is equal to

$$\Delta W = \Delta E - \Delta(TS) + p_0 \Delta V = \Delta F + p_0 \Delta V. \quad (57.2)$$

It should be stressed that formula (57.2) shows that the work ΔW represents the work done on the subsystem by an external source of work (but not by the medium).

For a small isothermal change in the volume ΔV the free energy in formula (57.2) can be expanded in a series in powers of ΔV , and rewritten in the form

$$\begin{aligned} \Delta W &= p_0 \Delta V + \left(\frac{\partial F}{\partial V} \right)_T \Delta V + \left(\frac{\partial^2 F}{\partial V^2} \right)_T \frac{(\Delta V)^2}{2} + \dots \approx \\ &\approx p_0 \Delta V - p \Delta V - \left(\frac{\partial p}{\partial V} \right)_T \frac{(\Delta V)^2}{2}. \end{aligned} \quad (57.3)$$

Since the process can be considered as quasi-static, in the process of fluctua-

tion the equilibrium pressure in the subsystem can be assumed to be equal to the pressure in the medium. Hence we find finally

$$\Delta W = - \left(\frac{\partial p}{\partial V} \right)_T \frac{(\Delta V)^2}{2}. \quad (57.4)$$

Substituting (57.4) into (55.5), we find the probability for the volume V to lie between V and $V + dV$:

$$dw = \text{const} \cdot \exp \left[\left(\frac{\partial p}{\partial V} \right)_T \frac{\Delta V^2}{2kT} \right] dV. \quad (57.5)$$

The constant is found from the normalization condition:

$$\text{const} \cdot \int_{-\infty}^{\infty} \exp \left[\left(\frac{\partial p}{\partial V} \right)_T \frac{(\Delta V)^2}{2kT} \right] dV = 1. \quad (57.6)$$

From formulae (57.5) and (57.6) it follows that the derivative $(\partial p / \partial V)_T$ must be negative. If this condition turned out to be unfulfilled, the probability of fluctuation would not decrease but increase with its magnitude. In such a substance there would occur volume fluctuations as a result of which the volume of the system would indefinitely increase or decrease down to zero. The substance would be in an unstable state. Thus, the condition of stability of the states of a homogeneous substance is given by the formula

$$\left(\frac{\partial p}{\partial V} \right)_T < 0. \quad (57.7)$$

If condition (57.7) is fulfilled, the integral (57.6) can easily be calculated. Then

$$\text{const} = \left(\frac{2\pi kT}{|(\partial p / \partial V)_T|} \right)^{-\frac{1}{2}}.$$

The normalized probability distribution of isothermal volume fluctuations has the form

$$dw = \left(\frac{|(\partial p / \partial V)_T|}{2\pi kT} \right)^{\frac{1}{2}} \exp \left[- \left| \left(\frac{\partial p}{\partial V} \right)_T \right| \frac{(V - V_0)^2}{2kT} \right] dV. \quad (57.8)$$

By means of the probability distribution (57.8) we find the mean square fluctuation of the volume $(\Delta V)^2 = \overline{(V - V_0)^2}$. Obviously, we have

$$\begin{aligned} \overline{(\Delta V)^2} &= \overline{(V - V_0)^2} = \\ &= \left(\frac{|\partial p / \partial V|_T}{2\pi kT} \right)^{\frac{1}{2}} \int_{-\infty}^{\infty} (V - V_0)^2 \exp \left(- \left| \left(\frac{\partial p}{\partial V} \right)_T \right| \frac{(V - V_0)^2}{2kT} \right) dV = \\ &= \frac{kT}{|\partial p / \partial V|_T}. \end{aligned} \quad (57.9)$$

Introducing the value of $\overline{(\Delta V)^2}$ into the distribution (57.8), one can rewrite it in a more compact form:

$$dw = [2\pi \overline{(\Delta V)^2}]^{-\frac{1}{2}} \exp \left[- \frac{(V - V_0)^2}{2 \overline{(\Delta V)^2}} \right] dV. \quad (57.10)$$

It follows from formulae (57.9) and (57.10) that the scale and probability of fluctuations increase with increasing temperature of the substance, as well as with increasing isothermal compressibility.

We apply formula (57.9) to the case of an ideal gas:

$$\overline{(\Delta V)^2} = \frac{kT}{|\partial p / \partial V|_T} = \frac{V^2 kT}{NkT} = \frac{V^2}{N}. \quad (57.11)$$

In what follows we shall be interested in the value of the mean square fluctuation of the density $\rho = v_0^{-1} = mV^{-1}$ (where m is the mass contained in the volume V in which the fluctuation takes place). We have

$$\overline{(\Delta \rho)^2} = m^2 \overline{\left(\Delta \frac{1}{V} \right)^2} = \frac{m^2}{V^4} \overline{(\Delta V)^2} = \frac{m^2}{V^2} \frac{kT}{V^2 |\partial p / \partial V|_T} = \frac{\rho^2}{V} kT \gamma_T,$$

where γ_T is the isothermal compressibility. The relative density fluctuation in the volume V is equal to

$$\overline{\left(\frac{\Delta \rho}{\rho} \right)^2} = \frac{kT \gamma_T}{V}. \quad (57.12)$$

We shall also find the fluctuation of the number of particles confined in a given volume. The quantity $(\Delta V)^2$ represents the mean square fluctuation of

the volume V of the system which contains N particles. The fluctuation of the volume per particle V/N is equal to

$$\overline{\left(\Delta \frac{V}{N}\right)^2} = \frac{kT}{N^2 |(\partial p / \partial V)_T|}.$$

Assuming the volume V to be fixed, we find the fluctuation of the number of particles in this volume:

$$\overline{(\Delta N)^2} = \frac{N^2 kT}{V^2 |(\partial p / \partial V)_T|}. \quad (57.13)$$

In particular, for an ideal gas

$$\overline{(\Delta N)^2} = N. \quad (57.14)$$

The independence of the fluctuation of the number of particles in a given volume on the temperature of the ideal gas is associated with the fact that in an ideal gas the motion of each particle is independent of the motion of other particles. With increasing temperature in an ideal gas only the mean square velocity increases, but the character of the motion does not change.

Formulae (57.8) and (57.9) lose their meaning in the case where the isothermal compressibility becomes infinite (and the derivative $(\partial p / \partial V)_T$ reduces to zero). A formal application of formula (57.9) leads to an absurd result: an infinitely large fluctuation of the volume. In reality, however, when the derivative $(\partial p / \partial V)_T$ reduces to zero the expression (57.4) for the work ΔW , on which the derivation of formula (57.9) is based, changes. The expansion of (57.3) must be continued, so that instead of (57.3) we must write

$$\begin{aligned} \Delta W &\approx p_0 \Delta V + \left(\frac{\partial F}{\partial V}\right)_T \Delta V + \left(\frac{\partial^2 F}{\partial V^2}\right)_T \frac{(\Delta V)^2}{2} + \left(\frac{\partial^3 F}{\partial V^3}\right)_T \frac{(\Delta V)^3}{6} + \dots \approx \\ &\approx p_0 \Delta V - p \Delta V - \left(\frac{\partial p}{\partial V}\right)_T \frac{(\Delta V)^2}{2} - \left(\frac{\partial^2 p}{\partial V^2}\right)_T \frac{(\Delta V)^3}{6}. \end{aligned} \quad (57.15)$$

We assume that the second derivative $(\partial^2 p / \partial V^2)_T$ is different from zero. Then, dropping infinitesimal quantities of higher order in (57.15), one can write

$$\Delta W = - \left(\frac{\partial^2 p}{\partial V^2}\right)_T \frac{(\Delta V)^3}{6}. \quad (57.16)$$

Substituting (57.16) into the normalization condition (57.6), we see that it cannot be satisfied for any value of the constant $(\partial^2 p / \partial V^2)_T$. This means that the assumption that the condition $(\partial^2 p / \partial V^2)_T \neq 0$ can be fulfilled for $(\partial p / \partial V)_T = 0$ leads to a contradiction.

Hence it is seen that if $(\partial p / \partial V)_T = 0$, then at the same time the condition

$$\left(\frac{\partial^2 p}{\partial V^2} \right)_T = 0. \quad (57.17)$$

must be fulfilled.

These two conditions determine the position of the critical point (see §64).

At the critical point the probability of a density fluctuation turns out to be considerably higher than in the ordinary state of substances, because here the work done in an isothermal change in the volume is very small.

It should, however, be stressed that for finding a quantitative expression for the probability distribution of fluctuations at the critical point the use of formula (55.5) with the substitution of the expansion (57.15) in it would be incorrect.

In the critical state of a substance its compressibility is so large that small forces have a great effect. Because of this, fluctuations here are not only large but, what is most important, lose their local character. This means that it is senseless to make a statement about the volume fluctuation at a given point of the substance*.

Now consider the fluctuations of the temperature of a subsystem at constant volume. The work which should be done on the subsystem in order to bring it from an equilibrium state with temperature T_0 into a non-equilibrium state with temperature T is equal to

$$\Delta W = \Delta E - T_0 \Delta S.$$

We expand the energy change ΔE in a series in powers of ΔS and confine ourselves to the first terms of the expansion. Since the energy is a potential with respect to the entropy and volume, we have

* The theory of fluctuations for a substance at the critical point cannot be expounded in this book. The reader is referred to the following book: L.D.Landau and E.M.Lifshitz, *Course of theoretical physics, Vol. 5: Statistical physics* (Pergamon Press, London, 1958).

$$\Delta E \approx \left(\frac{\partial E}{\partial S} \right)_V (\Delta S) + \left(\frac{\partial^2 E}{\partial S^2} \right)_V \frac{(\Delta S)^2}{2} = T_0 \Delta S + \left(\frac{\partial T}{\partial S} \right)_V \frac{(\Delta S)^2}{2}.$$

But

$$\Delta S = \left(\frac{\partial S}{\partial T} \right)_V \Delta T.$$

Hence, finally,

$$\begin{aligned} \Delta W &= T_0 \Delta S + \left(\frac{\partial T}{\partial S} \right)_V \frac{(\Delta S)^2}{2} - T_0 \Delta S = \\ &= \left(\frac{\partial T}{\partial S} \right)_V \left(\frac{\partial S}{\partial T} \right)_V \frac{(\Delta T)^2}{2} = \left(\frac{\partial S}{\partial T} \right)_V \frac{(\Delta T)^2}{2} = \frac{C_V}{2T_0} (\Delta T)^2. \end{aligned}$$

The probability that the temperature of the subsystem will undergo a fluctuation and that its temperature will lie between T and $T + dT$ is equal to

$$dw = \text{const} \cdot \exp \left(-\frac{C_V(T-T_0)^2}{2kT_0} \right) dT. \quad (57.18)$$

Normalizing the distribution (57.18), we find

$$dw = \left(\frac{C_V}{2\pi kT_0^2} \right)^{\frac{1}{2}} \exp \left(-\frac{C_V(T-T_0)^2}{2kT_0^2} \right) dT. \quad (57.19)$$

From the distribution (57.19) it follows that the heat capacity of a homogeneous substance at constant volume must be an essentially positive quantity. Otherwise the substance would be in an unstable state. Thus, in addition to (57.7) we obtain the second condition of stability of the states of a homogeneous substance:

$$C_V > 0. \quad (57.20)$$

If the heat capacity of the body were negative, then the body could be heated taking heat away from it. In other words, a perpetual motion machine of the second kind could be constructed.

It can be shown that the fluctuations of the volume and temperature are independent. Without dwelling on a strict proof of this statement, we note

only that it also follows from general physical reasoning. The state of a homogeneous body is completely determined by three thermodynamic parameters which are interconnected by one relation: the equation of state. Hence changes in two thermodynamic parameters in a homogeneous body can always take place independently of each other. In a homogeneous substance the conditions (57.7) and (57.20) are sufficient for the stability of the states of the system. It should be noted that the necessary conditions for stability are the constancy of the temperature and pressure in a homogeneous system.

In conclusion we note that the conditions of stability which we have obtained are not necessarily fulfilled in a non-homogeneous system, for example in a system placed in a field of force, or in a system consisting of several phases. In this case the state of the system depends, for example, on the strength of the external field as well as the parameters p , T , S and V , and other quantities. Hence the expressions for the work done in a fluctuation and the conditions of stability will be changed.

§58. Effect of fluctuations on the sensitivity of measuring devices

Fluctuations play an important role in the operation of modern sensitive instruments: balances, galvanometers, etc. The sensitivity of such devices is so high that they allow one to observe phenomena of the same scale as fluctuations caused by the thermal motion of molecules in the device itself. This leads to an important consequence: in a single direct measurement of a physical quantity whose value is smaller than the fluctuations of the device itself, the latter records its proper thermal motion (background) and not the quantity to be measured. In this sense it is said that the thermal motion imposes a limit of sensitivity upon a given form of the device (for a single measurement).

A further increase of the sensitivity and the measurements of quantities which lie below the thermal motion background are associated with carrying out repeated measurements (or with a change in the construction of the device).

Indeed, if a device records only its proper thermal motion, then its mean deflection will be equal to zero. But if an external action is superposed on the background, then the device will fluctuate about a new position, and its mean deflection will be different from zero. The larger the number of measurements carried out, i.e. the longer the observation time, the lower the values of a physical quantity (lying below the background) which can be measured.

We shall illustrate this by the analysis of several simple examples.

A suspended small mirror. One of the simplest and most sensitive devices is a light small mirror suspended on a fine filament, usually of quartz. The sensitivity of the device is determined by the possibility of recording very small angles of rotation of the mirror on the filament. The limit of sensitivity, i.e. the smallest angles of rotation which can be recorded in single measurements are determined by the fact that they must be larger than the oscillations of the mirror caused by the thermal motion of the molecules of the mirror and filament. This thermal motion leads to random rotations of the suspended mirror through angles whose order of magnitude is determined by the value of the mean square angle of rotation. We shall calculate this quantity.

In order that the mirror "randomly", i.e. under the action of the molecular thermal motion, be deflected from the equilibrium position $\varphi = 0$ through a certain angle φ , it is necessary that work be done against the elastic forces of the filament. This work is produced on account of the energy of the thermal motion. The angle φ plays the role of the parameter determining the deflection of the system from the equilibrium position. The probability of the deflection of the system from the equilibrium position $\varphi = 0$ through an angle φ is determined by formula (55.6) in which the potential energy of the torsion of the filament will stand for the potential energy. For small angles of rotation

$$u(\varphi) = \frac{1}{2}a\varphi^2,$$

where $a = \pi^2 r^2 G / 2l$ (here r is the radius of the filament, l is its length, and G is the shear modulus of the material of the filament).

Thus,

$$dw = \text{const} \cdot e^{-a\varphi^2/2kT} d\varphi = \frac{e^{-a\varphi^2/2kT} d\varphi}{\int_{-\infty}^{\infty} e^{-a\varphi^2/2kT} d\varphi}. \quad (58.1)$$

Here the constant is determined from the normalization condition.

The mean square angle of the deflection is equal to

$$\overline{\varphi^2} = \frac{\int_{-\infty}^{\infty} \varphi^2 e^{-a\varphi^2/2kT} d\varphi}{\int_{-\infty}^{\infty} e^{-a\varphi^2/2kT} d\varphi} = \frac{kT}{a}.$$

This result has a simple meaning: the mean potential energy of our system with one degree of freedom is equal to

$$\bar{u} = \overline{\frac{1}{2}a\varphi^2} = \frac{1}{2}kT \quad (58.2)$$

in accordance with the law of equipartition. For $T = 300$ K and $a = 10^{-6}$ erg (such a value of a is possessed by very fine quartz filaments) we have $(\varphi^2)^{\frac{1}{2}} = 2 \times 10^{-4}$. This quantity determines the angle through which the small mirror rotates, one the average, "by itself". If a quantity measured from a deflection of the mirror causes a rotation through a smaller angle, then in a single measurement it is the proper deflection due to thermal motion which is recorded.

It is clear, however, that in the absence of a systematic deflection force the mean deflection of the small mirror will be equal to zero, while in the presence of such a force the mirror will undergo oscillations about a displaced position of equilibrium. Carrying out repeated measurements of the oscillations of the small mirror one can find the mean position about which the oscillations take place. By this means one can determine a quantity whose values lie below the thermal background or sensitivity for a single measurement.

Spring-balance. Completely analogous results can be obtained for the spring-balance. The pressure fluctuations of the air surrounding it and the thermal motion of the mechanism of the balance will lead to random changes in the load of the balance. This change in the load will be compensated for by a quasi-elastic force $\kappa \Delta x$. The change in the potential energy of the system for a displacement by Δx is equal to

$$u = \frac{1}{2}\kappa(\Delta x)^2.$$

The mean potential energy, according to the law of equipartition, is equal to $\frac{1}{2}kT$. Hence the mean change in the length of the spring is equal to

$$[(\Delta x)^2]^{\frac{1}{2}} = (kT/\kappa)^{\frac{1}{2}}. \quad (58.3)$$

The measurement of a mass m by the balance is possible if the extension of the spring caused by it is larger than the length fluctuation $[(\Delta x)^2]^{\frac{1}{2}}$ of the spring. The extension of the spring by a load m is equal to $\Delta x = mg/\kappa$. Hence the limiting small mass which can be found in a single measurement is equal to

$$m \approx \kappa g^{-1} [(\Delta x)^2]^{\frac{1}{2}} = (kT\kappa)^{\frac{1}{2}} g^{-1}.$$

Gas thermometer. Suppose that we are measuring a temperature using a gas thermometer filled with an ideal gas. The temperature measured by the ther-

momenter will not remain constant but will continuously undergo fluctuations just as other thermodynamic quantities.

In an ideal gas the fluctuation of the temperature can easily be expressed in terms of the fluctuation of the volume. From the Clapeyron equation it follows that

$$\Delta T = \frac{p \Delta V}{Nk} = \frac{T}{V} \Delta V,$$

where ΔT and ΔV denote small changes in the temperature and volume. If small changes in the volume are understood to be changes due to fluctuations, then one can write

$$\Delta V = [\overline{(\Delta V)^2}]^{\frac{1}{2}} = \left(\frac{kT}{(-\partial p / \partial V)_T} \right)^{\frac{1}{2}} = \frac{V}{N^{\frac{1}{2}}},$$

so that

$$\Delta T = [\overline{(\Delta T)^2}]^{\frac{1}{2}} = TV^{-1} \Delta V = TN^{-\frac{1}{2}}.$$

Changes in the temperature which are smaller than ΔT cannot be measured by means of a gas thermometer. If the thermometer contains altogether 10^{-4} mole of a gas (i.e. if its volume is 0.02 litre), then $N = 6 \times 10^{23} \times 10^{-4} = 6 \times 10^{19}$, so that the minimal measurable change in the temperature is

$$\Delta T \approx 10^{-10} T.$$

This is so small that all really measurable changes in the temperature are extremely large in comparison with the limit of sensitivity.

Thus, the sensitivity of a gas thermometer is not in practice limited by fluctuations in the temperature. The examples given show that the effect of fluctuations on the sensitivity of devices varies widely depending on the character of the device.

Systems with a Variable Number of Particles

§59. The Gibbs grand canonical distribution

In considering in §13 the interaction of a subsystem with the bodies surrounding it (the reservoir) we have assumed that this interaction consists only of an energy exchange. In reality, however, the interaction of a subsystem with its surroundings amounts not only to an exchange of energy but very often also involves an exchange of particles. In the process of interaction the subsystem exchanges particles with the medium surrounding it. Particles which are going out and coming in carry energy with them, so that the energy exchange and particle exchange take place simultaneously. In this case not only the energy but also the number of particles in the system is variable. In order to characterize the state of such a system it is insufficient to indicate the total energy of the system; it is also necessary to indicate how many particles are contained in the system. Owing to the interaction with the surroundings, the subsystem singled out can be in different quantum states which differ by the number of particles contained in the system. Before passing over to the derivation of the statistical distribution for this case, we shall present certain examples of subsystems with a variable number of particles.

We assume that our subsystem represents a macroscopic drop or crystal

which is in equilibrium with the vapour or melt respectively. The latter play the role of the surroundings (the reservoir). Molecules from the surface of the liquid go to the vapour, while molecules from the vapour condense on the surface of the liquid. The same occurs with molecules on the surface of the crystal. If there is no systematic transfer of particles from the vapour to the liquid or vice versa, then in the system an equilibrium state will be established in which the number of particles going in the two directions is equal.

As another example of a system with a variable number of particles we consider one in which an equilibrium chemical reaction occurs. In the course of the chemical reaction the number of particles in the subsystem which has been selected changes (for example, the molecules of a compound AB): it decreases owing to the decomposition $AB \rightarrow A+B$ and increases owing to the synthesis $A+B \rightarrow AB$.

In an equilibrium state a continuous exchange of energy and particles takes place between the subsystem and the reservoir. Equilibrium conditions for the energy exchange were found to be the equality of the temperatures and pressures. An additional equilibrium condition for the exchange of particles will now be found.

We consider the derivation of the statistical distribution of a system with a variable number of particles, i.e. the distribution of the probabilities w_{in} that the subsystem will be found in the i th state and contain n particles. Finding the statistical distribution in this case differs from the case considered in § 16 only by the fact that the number of states of the subsystem with given energy $\Omega(\epsilon_i)$ must be replaced by the number of states with given energy and given number of particles $\Omega(\epsilon_i, n)$. Correspondingly, the number of states of the reservoir will be $\Omega_0(E_0, N_0)$. The sum of the number of particles in the subsystem and reservoir remains constant:

$$N = n + N_0 = \text{const.}$$

Then instead of formula (16.5) we obtain

$$w_{in} \approx \Omega_0(E - \epsilon_i, N - n) \Omega(\epsilon_i, n),$$

and, instead of (16.7),

$$\Omega_0(E - \epsilon_i, N - n) = \exp \{ \sigma(E - \epsilon_i, N - n) \}. \quad (59.1)$$

Since the dimensions of the subsystem are small, its energy and the number of particles contained in it are small in comparison with the energy and the

number of particles in the entire closed system: $\epsilon_i \ll E$ and $n \ll N$. Hence, as in §16, we can expand the function $\sigma(E - \epsilon_i, N - n)$ in a series in powers of ϵ_i and n and confine ourselves to the first terms of the expansion. This gives

$$w_{in} \approx \exp \left[\sigma(E, N) - \left(\frac{\partial \sigma}{\partial E_i} \right)_{\epsilon_i=0} \epsilon_i - \left(\frac{\partial \sigma}{\partial n} \right)_{n=0} n \right] \Omega(\epsilon_i, n), \quad (59.2)$$

or

$$w_{in} \approx \text{const} \cdot \exp \left[-\frac{\epsilon_i - \mu n}{\theta} \right] \Omega(\epsilon_i, n), \quad (59.3)$$

where the symbol "const" denotes the constant quantity $e^{\sigma(E, N)}$ which does not depend on ϵ_i and n ; θ denotes as before the statistical temperature $(\partial \epsilon_i / \partial \sigma)_{\epsilon_i=0}$, and

$$\frac{\mu}{\theta} = - \left(\frac{\partial \sigma}{\partial n} \right)_{n=0}. \quad (59.4)$$

The derivative in formula (59.4) is taken at a constant value of the energy and the external parameters. The molecular meaning of the quantity μ will be obtained in the next section.

It should be stressed that, in contrast to θ , μ can have any sign. Indeed, in formula (59.5) the summation is carried out over a finite number of particles, in contrast to the summation over an infinite number of levels in formula (16.12).

The value of the constant can be found from the normalization condition:

$$\sum_i \sum_n w_{in} = 1,$$

where the summation is carried out over all energy levels and all possible numbers of particles in the system. Obviously, we have

$$\text{const} \cdot \sum_i \sum_n \exp \left[\frac{\mu n - \epsilon_i}{\theta} \right] \Omega(\epsilon_i, n) = 1,$$

hence

$$\text{const} = \left[\sum_i \sum_n \exp \left(\frac{\mu n - \epsilon_i}{\theta} \right) \Omega(\epsilon_i, n) \right]^{-1}. \quad (59.5)$$

The probability distribution of the states of a system with a variable number of particles can finally be written in the form

$$w_{in} = \frac{\exp \left[\frac{\mu n - \epsilon_i}{\theta} \right] \Omega(\epsilon_i, n)}{\sum_i \sum_n \exp \left[\frac{\mu n - \epsilon_i}{\theta} \right] \Omega(\epsilon_i, n)}. \quad (59.6)$$

Formula (59.6) differs from formula (16.13) only by the fact that instead of one variable characterizing the state of the system, the energy, it contains two variables: the energy ϵ_i and the number of particles n in the system. We shall call the probability distribution (59.6) the grand canonical distribution.

We introduce the notation

$$\tilde{Z} = \exp \left[-\frac{\mu n}{\theta} \right] \sum_i \sum_n \exp \left[\frac{\mu n - \epsilon_i}{\theta} \right] \Omega(\epsilon_i, n).$$

For a constant number of particles in the system $n = \bar{n}$ the quantity \tilde{Z} is the same as the ordinary partition function.

The probability distribution (59.6) can be written by means of \tilde{Z} in the standard form:

$$w_{in} = \tilde{Z}^{-1} \exp [-\epsilon_i/\theta] \Omega(\epsilon_i, n).$$

The number of states of the system $\Omega(\epsilon_i, n)$ can (in the quasi-classical approximation) be expressed in terms of the volume $\Delta\Gamma$ of phase space according to formula (1.26):

$$\Omega(\epsilon_i, n) = \frac{\Delta\Gamma_n}{h^{3n}}, \quad (59.7)$$

where $\Delta\Gamma_n$ is the volume of the phase space of a system containing n particles. It is obvious that with a change in the number of particles the number of degrees of freedom $3n$ and the value of the phase volume also change:

$$\Delta\Gamma_n = \Delta q_1 \Delta q_2 \dots \Delta q_{3n} \Delta p_1 \Delta p_2 \dots \Delta p_{3n}. \quad (59.8)$$

Then for the probability that the system be in the energy state corresponding to the phase volume element $d\Gamma_n$ and contain n particles we obtain

$$dw_{in} = \frac{\exp[-\epsilon_i/\theta] d\Gamma_n}{\tilde{Z} h^{3n}}. \quad (59.9)$$

Knowing the probability distribution (59.6) or (59.9), one can find the mean values of all quantities characterizing a state of a system with a variable number of particles.

According to the general formula for obtaining mean values, we find the mean value of any quantity \bar{L} which depends on the state of the subsystem and on the number of particles:

$$\bar{L} = \frac{\sum_i \sum_n L \exp\left[\frac{\mu n - \epsilon_i}{\theta}\right] \Omega(\epsilon_i, n)}{\sum_i \sum_n \exp\left[\frac{\mu n - \epsilon_i}{\theta}\right] \Omega(\epsilon_i, n)}. \quad (59.10)$$

In particular, the mean value of the number of particles for an arbitrary value of the energy of the system is equal to

$$\bar{n} = \frac{\sum_i \sum_n n \exp\left[\frac{\mu n - \epsilon_i}{\theta}\right] \Omega(\epsilon_i, n)}{\sum_i \sum_n \exp\left[\frac{\mu n - \epsilon_i}{\theta}\right] \Omega(\epsilon_i, n)} = \theta \frac{\partial}{\partial \mu} \ln \sum_i \sum_n \exp\left[\frac{\mu n - \epsilon_i}{\theta}\right] \Omega(\epsilon_i, n). \quad (59.11)$$

For a system with a variable number of particles it is natural to call \tilde{Z} the grand partition function or the grand sum (or integral) over states.

It is convenient to introduce a quantity z which is called the activity and is by definition equal to $z = e^{\mu/\theta}$. By means of the activity, \tilde{Z} can be written in the form

$$\tilde{Z} = \sum_{n=1}^N z^n Z_n, \quad (59.12)$$

where Z_n is the statistical sum for n particles.

In the classical approximation we can write

$$\tilde{Z} = \sum e^{\mu n/\theta} Z_n \approx \sum e^{\mu n/\theta} e^{-F_n/\theta} \approx e^{(\mu \bar{n} - F)/\theta} \approx e^{(\mu N - F)/\theta}.$$

In the next paragraph it will be shown (see formula (60.4)) that $\mu N = \Phi$, where Φ is the Gibbs thermodynamic potential. Hence for \tilde{Z} we find

$$\tilde{Z} = e^{(\Phi - F)/\theta} = e^{pV/\theta}, \quad (59.13)$$

or

$$p = (\theta/V) \ln \tilde{Z}. \quad (59.14)$$

Analogously, (59.11) can be written in the form

$$\bar{n} = N = \theta \frac{\partial \ln \tilde{Z}}{\partial \mu} = \frac{\partial \ln \tilde{Z}}{\partial \ln z}. \quad (59.15)$$

We now have to consider the establishment of the physical meaning of the parameter μ .

In §17 the physical meaning of the quantity θ introduced formally has been elucidated, and it has been shown that it represents the statistical temperature. The condition of statistical equilibrium between quasi-independent subsystems which can interact weakly with each other and exchange energy was the equality of their temperatures. It is remarkable that the formally introduced quantity μ also turns out to have an important physical meaning which can be revealed by means of reasoning which is completely analogous to that of §17.

Consider a certain system which is in a state of statistical equilibrium. We single out from it two subsystems which are also in a state of statistical equilibrium and are weakly interacting with each other. This interaction consists of a mutual exchange of energies and particles between the two subsystems. For each of them one can write the probability distribution of states in the form

$$w_1 = A_1 \exp \left[\frac{\mu_1 n_1 - \epsilon_{i1}}{\theta} \right] \Omega_1 \quad \text{and} \quad w_2 = A_2 \exp \left[\frac{\mu_2 n_2 - \epsilon_{k2}}{\theta_2} \right] \Omega_2.$$

where the index 1 marks the quantities referring to the first subsystem, and the index 2 marks the quantities referring to the second subsystem.

Since the subsystems are quasi-independent, one can apply the theorem of multiplication of probabilities to them, and for the probability of finding simultaneously the first subsystem in the i th state and the second subsystem in the k th state one can write

$$w_{12} = w_1 w_2 = A_1 \exp \left[\frac{\mu_1 n_1 - \epsilon_{i1}}{\theta_1} \right] A_2 \exp \left[\frac{\mu_2 n_2 - \epsilon_{k2}}{\theta_2} \right] \Omega_1 \Omega_2. \quad (59.16)$$

On the other hand, the two subsystems together can be considered as one subsystem with an energy equal to the sum $\epsilon_{i1} + \epsilon_{k2}$ and with a number of particles equal to $n_1 + n_2$. Since this subsystem is an equilibrium in state, one can also write the grand statistical distribution for it in the form

$$w_{12} = A \exp \left[\frac{\mu(n_1 + n_2) - (\epsilon_{i1} + \epsilon_{k2})}{\theta} \right] \Omega. \quad (59.17)$$

If the subsystems are in equilibrium with each other, then their states should not change when an interaction is established between them. This means that the probability distribution of states in a system which is formed of two subsystems must remain unchanged. For this it is necessary that the expressions (59.16) and (59.17) should be identical. The latter condition requires, however, that the following equalities should be fulfilled:

$$\theta = \theta_1 = \theta_2, \quad (59.18)$$

$$\mu = \mu_1 = \mu_2. \quad (59.19)$$

The first of these is the well-known condition of equality of temperatures in all quasi-independent subsystems constituting an equilibrium system. This condition has been obtained in § 17 for subsystems whose interaction reduced to an energy exchange. The second equality is essentially new. It shows that the quantity μ which refers, as θ , to the reservoir (see § 17), in a state of statistical equilibrium must have the same value in all parts of the system.

In addition to the conditions of the constancy of the temperature and pressure, the constancy of μ is a necessary condition for statistical equilibrium in a system. The appearance of the additional condition of equilibrium is associated with the fact that we are now considering subsystems which can mutually exchange not only energy but also particles.

We shall call the quantity μ the chemical potential of the reservoir. In the case where the subsystem which we have singled out is itself a macroscopic system, the conditions of equilibrium allow one to refer μ to the system itself and not to the reservoir. Indeed, in an equilibrium state the chemical potentials of the reservoir and the macroscopic subsystem must be equal. However, it makes no sense to speak of the chemical potential of a microscopic subsystem, for example, a molecule. It should be noted that the same also holds for the statistical temperature θ . It also represents the temperature of the reservoir, but for a macroscopic subsystem it can be identified with the temperature of the latter. However, one cannot speak of the temperature of an individual molecule.

From the point of view of molecular concepts the condition (59.18) expresses the requirement that the amounts of energy which are given away and obtained by the subsystem should be equal to each other. The condition states that in the exchange of particles not only the numbers of particles coming in and going out of the subsystem must be equal to each other, but also the mean energies carried by the particles must be the same. If that were not so (for example, if only fast particles went out and only slow particles came in) then the equilibrium state would be violated.

§60. The basic thermodynamic equality and the calculation of chemical potentials

In order to elucidate the thermodynamic properties of a system with a variable number of particles it is first of all necessary to find the basic thermodynamic equality for such systems. The latter can be obtained most simply in the following way.

Since the quantities involved in the basic thermodynamic equality (24.5) — the energy, entropy and volume — possess additive properties, this equality can be written not only for the quantities E , σ and V but also for the specific values of these quantities related to unit mass or to one particle.

Let a system contain N particles. Then the energy, entropy and volume per particle can be written in the form E/N , σ/N and V/N . Writing the basic thermodynamic equality for the specific values related to one particle, we have

$$d\left(\frac{E}{N}\right) = \theta d\left(\frac{\sigma}{N}\right) - p d\left(\frac{V}{N}\right).$$

It is obvious that this equality will hold irrespective of the cause of the change in the specific value of the energy and the other quantities, i.e. irrespective of whether this change is due to a change in the quantities themselves or to a change in the number of particles in the system. Hence in the equality one can consider N to be a variable quantity and write

$$\frac{dE}{N} - \frac{E dN}{N^2} = \theta \frac{d\sigma}{N} - \theta \sigma \frac{dN}{N^2} + pV \frac{dN}{N^2} - \frac{p dV}{N},$$

hence

$$dE = \theta d\sigma - p dV + \left(\frac{E - \theta \sigma + pV}{N} \right) dN.$$

Setting

$$\mu = \frac{E - \theta \sigma + pV}{N}, \quad (60.1)$$

we obtain

$$dE = \theta d\sigma - p dV + \mu dN. \quad (60.2)$$

Whence it follows that the following equalities holds:

$$\mu = \left(\frac{\partial E}{\partial N} \right)_{\sigma, V}; \quad \frac{\mu}{\theta} = - \left(\frac{\partial \sigma}{\partial N} \right)_{E, V}. \quad (60.3)$$

The comparison of formulae (60.3) and (59.4) convinces us of the identity of the quantity μ determined by formula (60.1) and the chemical potential. The first of the equalities (60.1) shows that the chemical potential is equal to the derivative of the energy with respect to the number of particles. From the point of view of the practical calculation of chemical potentials the equality (60.1) is of special importance. It shows that the chemical potential μ represents the Gibbs thermodynamic potential related to one particle:

$$\mu = \frac{E - \theta \sigma + pV}{N} = \frac{\Phi(p, T)}{N}. \quad (60.4)$$

The latter justifies the term chemical potential. The chemical potential μ is

most conveniently expressed as a function of the pressure and temperature according to formula (60.4).

Formula (60.2) represents the basic thermodynamic equality written for a system with a variable number of particles.

We note in addition that, passing over in (60.2) from the energy to the free energy in the usual way, i.e. by subtracting the differential $d(TS)$ from both sides of (60.2), we can write

$$dF = -SdT - pdV + \mu dN, \quad (60.2')$$

whence

$$\mu = \left(\frac{\partial F}{\partial N} \right)_{T,V}. \quad (60.3')$$

The relations obtained can easily be generalized to the case of systems containing particles of different kinds. In what follows we shall consider the statistical properties of systems with a variable number of particles, and we shall need concrete expressions for chemical potentials. They can be obtained for gases and crystals.

By means of (37.11) we find for the chemical potential of an ideal monatomic gas

$$\mu = -\frac{5}{2}kT \ln kT + kT \ln p - kTj, \quad (60.5)$$

where the quantity j , which is often called the chemical constant, is equal to

$$j = \ln \left(\frac{2\pi m}{h^2} \right)^{\frac{3}{2}}.$$

For a diatomic gas one can analogously obtain

$$\mu = -\frac{7}{2}kT \ln kT + kT \ln p - kTj + kT \ln (1 - e^{-h\nu/kT}) + \epsilon_0, \quad (60.6)$$

where the chemical constant j and the zero point energy ϵ_0 are equal to

$$j = \ln \left[\left(\frac{2\pi m}{h^2} \right)^{\frac{3}{2}} \frac{8\pi^2 I}{h^2} \right]; \quad \epsilon_0 = \frac{1}{2}h\nu.$$

In the case of crystals, by the definition of the chemical potential we have

$$\mu = \frac{F + pV}{N} = \frac{F}{N} + pv,$$

where F is given by formulae (53.7) and (53.8), and v denotes the volume per particle in the crystal. At a low temperature $T \ll \theta_c$,

$$\mu = -\frac{\pi^4 kT}{5} \left(\frac{T}{\theta_c}\right)^3 + pv. \quad (60.7)$$

Analogously at a high temperature $T \gg \theta_c$,

$$\mu = -3kT \ln \frac{T}{\theta_c} - kT + pv. \quad (60.8)$$

The last formulae involve the product pv which is contained in the equation of state of the crystal expressed in terms of quantities which are difficult to measure. However, in view of the smallness of the volume v per particle, one can in most cases drop the small term pv .

In conclusion we shall make use of the value found for the chemical potential in order to write the Maxwell distribution in the form of the Gibbs distribution with a variable number of particles. For this we express μ not in terms of the pressure but in terms of the volume of the system. We obtain

$$\mu = kT \ln \left[\frac{N}{V} \left(\frac{h^2}{2\pi m} \right)^{\frac{3}{2}} \frac{1}{(kT)^{\frac{3}{2}}} \right], \quad (60.9)$$

whence

$$\frac{N}{V} \left(\frac{1}{2\pi m kT} \right)^{\frac{3}{2}} = \frac{1}{h^3} e^{\mu/kT}.$$

Substituting this into the Maxwell distribution (9.3), we obtain

$$dn = e^{(\mu - \epsilon)/kT} \frac{d\Gamma}{h^3}. \quad (60.10)$$

We note that the chemical potential of an ideal gas is a very large negative quantity.

§61. Conditions of phase equilibrium

One of the most important cases of statistical equilibrium in a system with a variable number of particles is that of phase equilibrium. Imagine a homogeneous quasi-closed macroscopic system, which is separated from other bodies by an interface and is in an equilibrium state. We shall call such a system a phase of the substance.

The concept of phase is a generalization and a more precise definition of the concept of state of aggregation. As an example of a phase one can point to a vapour which is in equilibrium with its condensate. In this case the exchange of energy and particles between the vapour and its surroundings, the reservoir, takes place through the interface vapour—liquid or vapour—solid. As other examples we mention: a crystal in equilibrium with its melt; a crystal modification in equilibrium with another; an electron gas in vacuum in equilibrium with an electron gas in a metal. In what follows we shall quote examples of other, more complex phase equilibria.

In all cases of phase equilibrium the existence of an interface between different phases is characteristic. In a state of statistical equilibrium the number of particles passing over from one phase into another, and the energy carried by them are exactly equal to the corresponding quantities in the reverse direction. If there were no interface sharply separating the phases, it would be senseless to speak of a particular quasi-closed subsystem as a phase. We shall illustrate this by an example. Imagine that two phases, representing uniform isotropic states of a substance, are in equilibrium, and are connected by a certain interface. Then we shall call the less dense phase the vapour or gas phase, and the more dense one the liquid phase. If however, we have a uniform system, then, as will be explained in detail in §64, the notion of liquid or gas is inapplicable to it: by changing the physical conditions of the system it can be transformed continuously from a state with a high density to one with a low density. The state with the high density cannot be called the liquid, and the state with the low density cannot be called the gas. They both represent cases of a uniform state of the substance. Thus, the presence of an interface between phases is the necessary condition for one to speak of the existence of phases and phase equilibria.

Let us write down the condition for statistical equilibrium between phases, confining ourselves in the beginning to two phases of a substance. Each of the phases can be considered as a quasi-closed subsystem, and the set can be considered as a closed system in a state of statistical equilibrium. Hence the conditions of equilibrium between two phases can be written in the form of (59.14) and (59.15):

$$T_1 = T_2, \quad (61.1)$$

$$\mu_1(p_1, T) = \mu_2(p_2, T). \quad (61.2)$$

In addition to these conditions it is necessary that the forces applied to the interface by the two equilibrium phases be equal to each other. Otherwise the interface between the phases would move and the equilibrium in the system would be violated.

It is convenient to relate the condition of chemical equilibrium to unit interface, replacing forces by pressures. Thus, in addition to the conditions of equality of temperatures and chemical potentials one has the condition of equal pressures in the two phases.

This simple reasoning can be strictly substantiated, by considering the condition for mechanical equilibrium, which is the requirement of minimum free energy of the closed system at $T = \text{const}$ and $\mu = \text{const}$. The condition of mechanical equilibrium in a system consisting of two phases at $T = \text{const}$ can be written in the form

$$dF = dF_1 + dF_2 = -p_1 dV_1 - p_2 dV_2 = 0.$$

Since the volume of the entire system remains unchanged,

$$dV_1 = -dV_2$$

and

$$p_1 = p_2. \quad (61.3)$$

Thus, the pressures in the two phases must be equal to each other. Taking into account the conditions (61.1) and (61.3), formula (61.2) can be written as

$$\mu_1(p, T) = \mu_2(p, T). \quad (61.4)$$

Since in a state of equilibrium T and p have equal values in the two phases, on the basis of eq. (61.4) one of these quantities can be expressed in terms of the other. As a result of this we obtain the equation

$$p = p(T) \quad (61.5)$$

for the dependence of the equilibrium pressure on the equilibrium temperature. Eq. (61.5) represents a certain curve in the (p, T) plane, called the phase equilibrium curve. All points of this curve correspond to the contact of equilibrium phases. At pressures which are larger and smaller than the equilibrium pressure at given temperature, one of the phases, the one which has a lower thermodynamical potential, is stable. If, for example, one of the phases is a liquid, and the second is its vapour, then the region in the (p, T) -plane which lies above the curve corresponds to the liquid phase, while the region lying below the curve corresponds to the gas phase. The equilibrium liquid-gas phase transition takes place along the curve.

Correspondingly, in the crystal-melt equilibrium the region above the phase equilibrium curve corresponds to the crystal phase, melting points lie on the curve, and the stable phase below the curve is liquid. In a phase transition the liberation or absorption of latent heat takes place. The latent heat for the transition of a molecule from one phase into the other is equal to (since the process is a reversible and equilibrium process)

$$l = \int T ds ,$$

where s is the entropy related to one molecule. Since in the phase transition the temperature is constant, it can be taken out of the integral sign and one can write

$$l = T\Delta s = T(s_2 - s_1) .$$

Thus, the latent heat of a phase transition is equal to the difference between the entropies multiplied by the temperature of the transformation. The latent heat is taken to be positive if heat is absorbed in the phase transition. Latent heat which is liberated is taken to be negative.

§62. The equation of the phase equilibrium curve. Equilibrium between the vapour and the condensed phase

The dependence of the chemical potential on the temperature and pressure is known for only a few simple systems. In most cases the concrete form of the function $\mu(p, T)$ is unknown. Hence the equation of the equilibrium curve (61.5) cannot be written in an explicit form. However, it turns out that the differential equation of the equilibrium curve has a simpler form and contains only quantities which can easily be measured.

To obtain the differential equation of the equilibrium curve we differentiate the condition (61.4). We have

$$d\mu_1 = d\mu_2 \quad (62.1)$$

or

$$\frac{\partial\mu_1}{\partial p} dp + \frac{\partial\mu_1}{\partial T} dT = \frac{\partial\mu_2}{\partial p} dp + \frac{\partial\mu_2}{\partial T} dT. \quad (62.2)$$

From formula (62.2) we find the slope of the equilibrium curve

$$\frac{dp}{dT} = \frac{\partial\mu_2/\partial T - \partial\mu_1/\partial T}{\partial\mu_1/\partial p - \partial\mu_2/\partial p}. \quad (62.3)$$

Eq. (62.3) is just the differential equation of the equilibrium curve sought. In order to bring it into final form it is necessary to express the quantities appearing in it in terms of those measured directly. According to (29.10) and (29.11) we have

$$S = -\frac{\partial G}{\partial T}, \quad V = \frac{\partial G}{\partial p}.$$

Hence

$$\frac{\partial\mu}{\partial T} = -\frac{S}{N} = -s, \quad \frac{\partial\mu}{\partial p} = \frac{V}{N}, \quad (62.4)$$

where N is the number of particles. Substituting this into (62.3), we find

$$\frac{dp}{dT} = \frac{S_1 - S_2}{V_1 - V_2}. \quad (62.5)$$

Replacing the difference between the entropies by the heat of transformation l , we obtain

$$\frac{dp}{dT} = \frac{lN}{(V_2 - V_1)T}. \quad (62.6)$$

Formula (62.6) is usually referred to one mole of the gas phase.

Denoting the latent heat of the phase transition of one mole of a substance lN by L , and the change in the molar volume by ΔV , we find finally

$$\frac{dp}{dT} = \frac{L}{T\Delta V}. \quad (62.7)$$

Formula (62.6) is called the Clapeyron–Clausius equation. The Clapeyron–Clausius formula relates the change in the equilibrium pressure p for an infinitesimal change in the equilibrium temperature T to directly measured quantities. We shall discuss it for concrete cases of phase equilibria in following sections.

It is easily seen that if a phase transition takes place as the temperature increases, then the latent heat is always absorbed, i.e. $L > 0$. Indeed,

$$L = T(S_2 - S_1) = NT \left(\frac{\partial \mu_1}{\partial T} - \frac{\partial \mu_2}{\partial T} \right). \quad (62.8)$$

The character of the temperature variation of chemical potentials in a phase transition taking place with an increase in the temperature is shown in fig. III.36. Up to point 1 the stable phase is the first one, whose chemical potential μ_1 is smaller than that of the second phase μ_2 . After point 1 the situation is reversed. At point 1 phase equilibrium occurs. At it the chemical potentials of the two phases are equal to each other. Its ordinate represents the temperature of the phase transition (at a given pressure). From fig. III.36 it is seen that at point 1 the slope of curve μ_1 must be larger than that of curve μ_2 . Otherwise above this point μ_1 will not become larger than μ_2 .

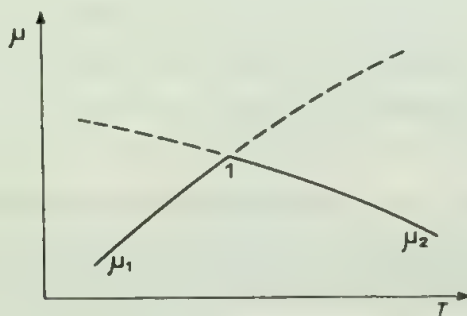


Fig. III.36

Hence at point 1 we have

$$\frac{\partial \mu_1}{\partial T} > \frac{\partial \mu_2}{\partial T}.$$

Then it follows from formula (62.8) that if the phase transition takes place with an increase in temperature its latent heat is always positive. The numerical value of the latter cannot be found theoretically, since it is expressed in terms of the entropies of the phases, and the explicit form of these functions is unknown in most cases.

The theorem proved above allows one to establish the sign of the temperature coefficient of the equilibrium pressure dp/dT for different phase transitions. If a phase transition takes place with an increase in the temperature (melting, boiling, sublimation), so that L is positive, then, according to the Clapeyron–Clausius formula, the sign of dp/dT is determined by the sign of the quantity ΔV , the change in volume in the phase transition. In evaporation and sublimation the volume of the phase increases sharply, so that $\Delta V > 0$ always. Hence for these phase transitions dp/dT is also positive, i.e. the equilibrium pressure increases with increasing temperature or, conversely, the equilibrium temperature increases with increasing pressure. As the pressure is lowered the temperatures of the boiling point and of the sublimation point decrease. Such a relation between the equilibrium pressure and the equilibrium temperature is in agreement with well-known experimental facts (increase in the boiling point in high-pressure boilers, decrease in the boiling point with the height, and so on).

In melting two cases are encountered: when ΔV is positive, so that the density of the liquid phase is smaller than that of the solid phase, and when ΔV is negative, so that the liquid phase is more dense. For bodies of the first type $dp/dT > 0$, so that the melting point increases with increasing pressure.

The number of bodies which are more dense in the liquid phase is relatively small. Examples of such bodies are water, cast iron, bismuth, and a number of alloys. For these $dp/dT < 0$, i.e. the melting point decreases with increasing pressure. This feature of the melting of ice and other substance is well known.

It is interesting to note that in the vicinity of absolute zero the temperature coefficient dp/dT tends to zero, so that the equilibrium pressure at the melting point ceases to depend on temperature. Indeed, from the third law of thermodynamics it follows that in melting the change in the entropy $\Delta S \rightarrow 0$ as $T \rightarrow 0$. Consequently, the latent heat of melting reduces to zero, and with it, by virtue of (62.6), also dp/dT . Such behaviour of the dependence of

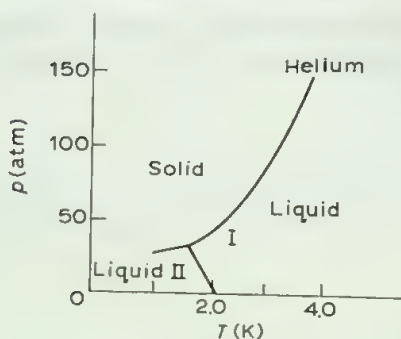


Fig. III.37

dp/dT on temperature does indeed take place for liquid helium II, which is a stable phase as $T \rightarrow 0$ at pressures below 30 atm. At pressures above ≈ 30 atm the stable phase is solid helium. The phase equilibrium curve (solid helium \rightleftharpoons liquid helium II) is almost horizontal. Its slope $dp/dT \rightarrow 0$ as $T \rightarrow 0$ (fig. III.37).

As we have already pointed out, the explicit form of the equilibrium curve cannot in general be found. If the dependence of the latent heat of transformation and of the change in the molar volume on temperature and pressure is known, then the Clapeyron–Clausius equation can be integrated. Then the dependence of the equilibrium pressure of the phase transition on temperature, i.e. the form of the equilibrium curve, can be found. The dependence of the quantities mentioned on temperature and pressure is usually complex, and the integration is carried out numerically. The situation is essentially simplified if one of the equilibrium phases is vapour, i.e. in the case of boiling or sublimation.

In the case of equilibrium between a condensed phase and vapour it can be assumed that the molar volume of the vapour is considerably larger than the molar volume of the condensed phase — a liquid or a crystal. Hence the change in the volume in the phase transition can be equated with the volume of the gas phase (related to the corresponding number of particles):

$$\Delta V = V_{\text{vap}} - V_{\text{cond.phase}} \approx V_{\text{vap}}.$$

In this case the Clapeyron–Clausius equation assumes the form

$$\frac{dp}{dT} = \frac{L}{TV_{\text{vap}}}. \quad (62.9)$$

If the vapour which is in equilibrium with the condensed phase is sufficiently rarefied, so that it can be considered as an ideal gas, then

$$V_{\text{vap}} = NkT/p. \quad (62.10)$$

It should be noted that this proposition is only fulfilled with sufficient accuracy at relatively low pressures. Substituting (62.10) into (62.9), we have

$$\frac{dp}{p} = \frac{L}{NkT^2} dT. \quad (62.11)$$

The dependence of L on the temperature can be found by means of a method which is completely analogous to that applied in deriving the Clapeyron–Clausius equation.

Differentiating L with respect to T , we have

$$\frac{dL}{dT} = T \left(\frac{\partial \Delta S}{\partial T} \right)_p + T \left(\frac{\partial \Delta S}{\partial p} \right)_T \frac{dp}{dT} + \Delta S$$

or

$$\frac{dL}{dT} = \Delta C_p - T \frac{\partial(\Delta V)}{\partial T} \frac{dp}{dT} + \Delta S,$$

or, taking into account (62.9) and (62.10),

$$\frac{dL}{dT} = \Delta C_p - \frac{T\Delta S}{V_{\text{vap}}} \frac{\partial V_{\text{vap}}}{\partial T} + \Delta S = \Delta C_p.$$

Thus, the latent heat of transformation at a temperature T is equal to

$$L = L_0 + \int_0^T (\Delta C_p) dT, \quad (62.12)$$

where L_0 is the latent heat at $T = 0$. This quantity represents the work which must be done at absolute zero in order to break the bonds existing between the molecules in the condensed phase and to transform them into non-interacting molecules. The latent heat of transformation from the condensed phase into the gas assumes a particularly clear meaning: it is equal to the

work done in order to overcome the bonds plus the energy which must be transferred to the system in order to compensate for the difference between the energies of thermal motion in the condensed phase and the gas.

Substituting formula (62.12) into (62.11), we have

$$\frac{dp}{p} = \frac{L_0 dT}{NkT^2} + \frac{dT}{NkT^2} \int_0^T (\Delta C_p) dT. \quad (62.13)$$

Integrating (62.13), we obtain

$$\ln p = -\frac{L_0}{NkT} + \int_0^T \frac{dT'}{NkT'^2} \int_0^{T'} (\Delta C_p) dT'' + i$$

or

$$p = \exp \left[-\frac{L_0}{NkT} + \int_0^T \frac{dT'}{NkT'^2} \int_0^{T'} (\Delta C_p) dT'' + i \right], \quad (62.14)$$

where i is a constant which is usually called the vapour pressure constant.

Formula (62.14) shows that the pressure of a saturated equilibrium vapour decreases rapidly with decreasing temperature.

In the case of evaporation the basic part of the latent heat of transformation usually corresponds to the first term of (62.12). Hence formula (62.14) is often approximately written in the form

$$p \sim \exp \left[-\frac{L_0}{NkT} + i \right]. \quad (62.15)$$

Formula (62.14) involves the unknown vapour pressure constant i , the latent heat of transformation at absolute zero, L_0 , and the difference between the heat capacities of the two equilibrium phases.

By means of statistical methods the value of all these quantities, except L_0 , can be obtained mathematically, provided the condensed phase is a crystal. The smallness of the pressure of the saturated vapour allows one to consider the vapour as an ideal gas and to make use of the chemical potential determined by formula (60.5).

Equating the chemical potential of the gas (for simplicity of the formulae a monatomic gas) and the crystal we obtain the equation of the sublimation curve. In this case we choose the zero energy in such a way that the energy of a motionless molecule of the gas is equal to zero. The energy of a crystal

molecule, measured from this level, is negative (since the molecule is bound in the crystal lattice) and will be denoted by $\epsilon_0 = -L/N$.

It is obvious that ϵ_0 is equal to the work which must be done at absolute zero to tear a molecule away from its neighbours in the crystal lattice and to bring it into the gas phase in a state in which it will also be at rest. Thus, $-\epsilon_0$ represents the heat of sublimation at absolute zero referred to one molecule and taken with the opposite sign. Equating μ_{gas} and μ_{cryst} , we find the conditions of equilibrium in a system (crystal \rightleftharpoons gas) at low temperature:

$$\ln p = \frac{5}{2} \ln kT + j - \frac{\epsilon_0}{kT} - \frac{\pi^4}{5} \left(\frac{T}{\theta_c} \right)^3,$$

or

$$p = (kT)^{\frac{5}{2}} \exp \left[-\frac{L_0}{NkT} - \frac{\pi^4}{5} \left(\frac{T}{\theta_c} \right)^3 \right] \exp j. \quad (62.16)$$

The basic term in (62.16) is the one which contains the latent heat of transformation at absolute zero.

Let us compare formula (62.16) with the general formula (62.14). For this we have to calculate the double integral in (62.14). Using for $C_{p \text{ cryst}}$ its value from formula (53.2), we have

$$\Delta C_p = -C_{p \text{ cryst}} + C_{p \text{ gas}} = \frac{5}{2} Nk - \frac{12\pi^4 Nk}{5} \left(\frac{T}{\theta_c} \right)^3$$

and

$$\begin{aligned} \int_0^T \frac{dT'}{NkT'^2} \int_0^{T'} \Delta C_p dT'' &= \\ &= \int_0^T \frac{dT'}{NkT'^2} \int_0^{T'} \left[\frac{5}{2} Nk - \frac{12\pi^4}{5} Nk \left(\frac{T''}{\theta_c} \right)^3 \right] dT'' = \\ &= \int_0^T \frac{dT}{T'^2} \left(\frac{5}{2} T' - \frac{3\pi^4 T'^4}{5\theta_c^3} \right) = \frac{5}{2} \ln T - \frac{\pi^4}{5} \left(\frac{T}{\theta_c} \right)^3. \end{aligned} \quad (62.17)$$

Substituting (62.17) into the general expression for the pressure (62.14), we obtain an expression which is the same as (62.16) provided the constant i

contained in (62.14) is assumed to be equal to the chemical constant j . Thus, the constant of the pressure of a saturated vapour can be calculated by means of statistical considerations.

Formula (62.16) is in good quantitative agreement with experiment.

The equilibrium curve at high temperatures can be obtained in exactly the same way.

Equating the chemical potentials (60.5) and (60.8), we find

$$\ln p = \frac{5}{2} \ln kT + j + \frac{\epsilon_0}{kT} - 3kT \ln \frac{T}{\theta_c} - 1. \quad (62.18)$$

The same result can also be obtained from the general formula (62.14).

It should be noted that, as is seen from (62.16), the pressure of a saturated vapour increases very rapidly with increasing temperature. If the characteristic temperature θ_c of the crystal is relatively large, so that the condition $T \gg \theta_c$ is fulfilled at high temperatures, then the corresponding density of the saturated vapour will be too large for the vapour to be considered as an ideal gas.

In this case it is necessary in formula (62.16) to make use of the chemical potential for a van der Waals gas. In practice empirical formulae for the vapour pressure curve are more often used.

It is useful to compare, for the example given, the practical potentialities of the thermodynamic and statistical methods.

We have obtained formula (62.14) which is of very general character and which establishes the equilibrium pressure of the vapour above any condensed phase by the thermodynamic method. However, this general formula involves quantities whose numerical value can be determined only from experimental data.

The expression for the vapour pressure has been obtained by statistical methods in the presence of strong restricting assumptions, but within this framework the quantitative values of all the quantities have been obtained and their molecular meaning has been elucidated.

§63. Theory of phase transitions

Up to now we have confined ourselves to thermodynamic reasoning, assuming as an experimental fact the existence of phases and the possibility of phase transitions.

We now have to discuss the phenomena of phase transitions from the

statistical point of view. A phase transition is always associated with the rupture of the continuity of certain thermodynamic quantities. In the example of a phase transition considered above, which is called a phase transition of the first kind, thermodynamic potentials remain continuous, whereas their entropy $S' = (\partial F / \partial T)_p$ and specific volume $v = V/N^{-1} = -N^{-1}(\partial F / \partial p)_T$ undergo a finite jump.

In addition to phase transitions of the first kind there are also the so-called phase transitions of the second kind, in which the second derivatives of the thermodynamic potentials — the heat capacity $C_p = T(\partial S / \partial T)_p$ and the coefficient of thermal expansion $\alpha = -V^{-1}(\partial V / \partial T)_p$ — undergo a break, changing discontinuously by amounts ΔC_p and $\Delta \alpha$.

We shall in what follows encounter numerous examples of phase transitions of the second kind (see Part IV §20 and §21). It should be stressed that the very existence of phase transitions appears to be rather unexpected from the point of view of statistical physics. One would think that the statistical sum (or integral) determines thermodynamic potentials as continuous functions of the parameters which characterize the state of the systems, for example, the temperature and volume.

Indeed, we can write the equation of state of a phase of a system with a given number of particles N occupying a volume V . It follows that from (59.14) that

$$p = \frac{kT}{V} \ln \tilde{Z} = \frac{kT}{V} \ln \sum z^n Z_n(V, T, n), \quad (63.1)$$

$$\frac{1}{v} = \frac{N}{V} = \frac{kT}{V} \frac{\partial \ln \tilde{Z}}{\partial \mu} = \frac{1}{V} \frac{\partial \ln \left[\sum z^n Z_n(V, T, n) \right]}{\partial \ln z} = \frac{\partial p}{\partial \ln z}. \quad (63.2)$$

Eliminating z from (63.1) and (63.2), one can find the function $f(p, V/N, T)$, i.e. the equation of state. If the law of interaction between molecules is taken in the form (46.2) and if the phase volume is assumed to have a finite fixed value V , then in Z_n one can carry out the integration over the momenta and write (in the classical approximation)

$$Z_n = \left(\frac{2\pi m kT}{h^2} \right)^{\frac{3}{2}} \frac{1}{n!} \int \exp \left[- \frac{\sum u(r_{ij})}{kT} \right] dV_1 \dots dV_n. \quad (63.3)$$

* In this section we follow the book by K. Huang, *Statistical mechanics* (John Wiley, New York, 1963). For the details of the theory we refer the reader to this book and to original studies by Yang and Lee, *Phys. Rev.* 87 (1952) 410.

It is obvious that all statistical integrals Z_n are essentially positive quantities which depend on the volume V and temperature T as parameters. We write the expression for \tilde{Z} in more detail:

$$\tilde{Z} = \sum_{n=1}^N Z_n z^n = Z_1 z + Z_2 z^2 + \dots + Z_N z^N. \quad (63.4)$$

The grand partition function \tilde{Z} is a polynomial of the n th power with respect to z with essentially positive coefficients. Hence \tilde{Z} is a monotonically increasing function of the activity z . By virtue of the continuity of functions, $Z_n(V, T, n)$ is also a continuous function of the temperature and volume (or the density $\rho = N/V$).

From the form of the function \tilde{Z} and formula (63.1) it is seen that $p(z)$ is a monotonically increasing function of the activity z , as is shown in fig. III.38a. The inverse specific volume $1/v$ is also, by virtue of (63.2), a monotonically increasing function of z (fig. III.38b). According to formula (57.7), the necessary condition for a stable existence of any phase is the requirement $(\partial p / \partial v)_T < 0$, i.e. the requirement of a monotonic decrease of the pressure with increasing volume per particle. Therefore the curve $p(v)$ has the form shown in fig. III.38c.

We see that at given temperature the pressure is a monotonic function of the volume and that there is no tendency for the appearance of discontinuities on the curve which expresses the equation of state of an arbitrary phase.

If, however, the function \tilde{Z} reduces to zero, then according to (63.1) and (63.2) the pressure p and the volume v will become indefinite, and our reasoning will lose its validity. Therefore it is necessary to discuss in more detail the behaviour of \tilde{Z} as a function of the activity z and the volume v .

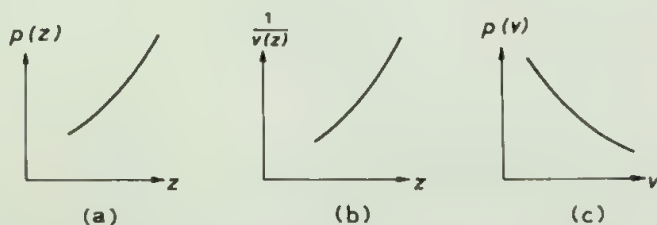


Fig. III.38

The very existence of phase transitions (discontinuities in thermodynamic quantities) is associated with the behaviour of the system as $\tilde{Z} \rightarrow 0$, where the functions p and v have singularities.

In order to find the values of z for which the grand partition function \tilde{Z} reduces to zero, it is necessary to know its explicit expression as a function of z and V . Finding the grand partition function for real systems appears, for the present, to be an impracticable problem.

It turns out, however, that in the limiting case where the number of particles N in the system and its volume V increase indefinitely ($N \rightarrow \infty$, $V \rightarrow \infty$) but in such a way that the specific volume v remains limited to $v \leq v_0$, it is possible to investigate the behaviour of $\tilde{Z}(z, v)$ without determining its explicit form. In this case it can be shown that from the behaviour of $\tilde{Z}(z)$ there follows the possibility of the existence of phase transitions.

We write the expression (63.4) in the form

$$\tilde{Z}(z, V, T) = \prod_{i=1}^N (z - z_i), \quad (63.5)$$

where z_i are the roots of the polynomial (63.4). Since all coefficients of the polynomial (63.4) are positive, the roots z_i cannot be positive. They are either negative, complex-conjugate quantities in pairs, or equal to zero.

Although only the positive and, in the last resort, zero values of z have a real meaning, from the mathematical point of view it is convenient to introduce into the treatment the complex values of z , and to consider the function of a complex variable $\tilde{Z}(z)$. Then we pass over to the limit in formulae (63.1), (63.2) and (63.5), writing

$$p = kT \lim_{V \rightarrow \infty} \left[\frac{1}{V} \ln \tilde{Z}(z, V, T) \right], \quad (63.6)$$

$$\frac{1}{v} = kT \lim_{V \rightarrow \infty} \left[\frac{1}{V} \frac{\partial \ln [\tilde{Z}(z, V, T)]}{\partial \ln z} \right], \quad (63.7)$$

$$\tilde{Z}(z, V, T) = \lim_{N \rightarrow \infty} \prod_{i=1}^N (z - z_i). \quad (63.8)$$

The number of roots z_i increases indefinitely, and they are distributed in a complex plane. A mathematical study carried out by Yang and Lee showed

that there are limits in formulae (63.6) and (63.7) and that for $V \rightarrow \infty$ the function $V^{-1} \ln \tilde{Z}(z)$ is an analytic function of z in certain regions R of the complex plane, including the real axis, which does not contain zero z_i .

This means that in a system with $V \rightarrow \infty$ for all values of z in these regions the pressure has no singularities (i.e. there are no phase transitions). In fig. III.39 these regions are denoted by R_1 and R_2 . Zero z_i are denoted by solid dots.

However, in contrast to a system with a finite value of N (a finite number of zeros), in a system with $N \rightarrow \infty$ and $V \rightarrow \infty$ the number of zero z_i is indefinitely large. Therefore, filling the plane of the complex variable, they can at certain points approach arbitrarily close to the real axis. Let z_0 be such a point on the real axis (fig. III.39). The zero of the function \tilde{Z} is located in an arbitrarily small region about the point z_0 . Since the pressure p is an analytic function at the point $z = z_0$, it must remain continuous (fig. III.40a). However, the inverse specific volume $1/v$ which, according to (63.2), is the derivative of the pressure can have a discontinuity at the point z_0 :

$$\Delta \left(\frac{1}{v} \right) = \Delta \frac{\partial \ln \tilde{Z}}{\partial \ln z}.$$

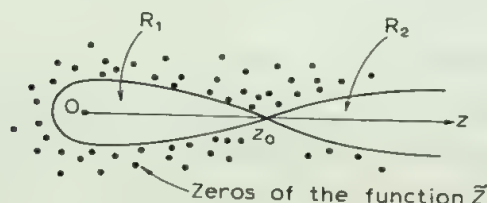


Fig. III.39

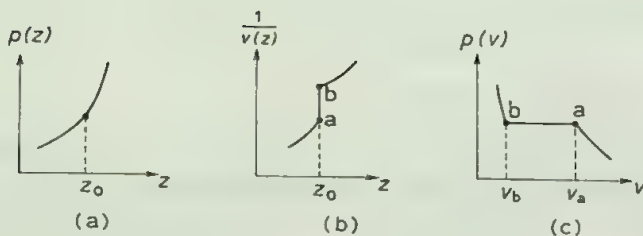


Fig. III.40

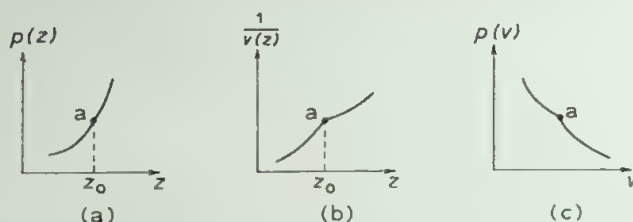


Fig. III.41

If such a discontinuity takes place as is shown in fig. III.40b, then the equation of state [the function $p(v)$] assumes the form shown in fig. III.40c. This is a typical curve of a phase transition of the first kind. When the specific volume changes from v_b to v_a , the pressure remains unchanged.

According to the above, a discontinuity in the derivative is possible but not obligatory. If, however, the pressure and its first derivative are continuous for $z = z_0$, then the second derivative of the pressure $\partial^2 p / \partial z^2$ can have a discontinuity. In this case a break arises on the curve $1/v(z)$, as is shown in fig. III.41b. Correspondingly, the equation of state will have the form shown in fig. III.41c. Such a curve is characteristic of phase transitions of the second kind.

It should be stressed that the theory which is discussed points to the possibility of phase transitions, but allows one to establish neither the position of the points of a phase transition nor the character of the transition itself. Also, the existence of isolated points z_0 is not proved. It is important therefore, that it has been possible to calculate the partition function in an explicit form for the simplest two-dimensional lattice consisting of particles which can be in two states (i.e. the Ising model). This calculation is cumbersome and cannot be presented here *. It turns out that in such a lattice one observes a phase transition of just the same type (with an isolated point z_0) as assumed above.

§64. Phase equilibrium curves

The phase equilibrium curve, i.e. the curve of the dependence of the equilibrium pressure p on the equilibrium temperature T in the (p, T) -plane,

* See L.D.Landau and E.M.Lifshitz, *Course of theoretical physics*, Vol. 5: *Statistical physics* (Pergamon Press, London, 1958).

has a different form for different phase equilibria. As we have already pointed out, the general form of the curve cannot be determined theoretically. One can only make some general observations apropos of this curve.

In considering the problem of the interrelation between a liquid and a gas (§48) we have already pointed out the absence of a fundamental difference between these states of matter. The existing qualitative differences between a liquid and a gas are associated with different roles of the interaction between atoms. As the temperature of a gas decreases or the density increases the mean distance between the atoms decreases. This corresponds to a decrease in the mean free path and to a relative increase in the mean energy of interaction (in comparison with kT). Under certain conditions the thermodynamic potential of a system of widely spaced freely moving particles (a gas) turns out to be higher than that of a system in which the distances between the molecules are small (a liquid). At this moment a phase transition (condensation) takes place. The chaotic free motion of molecules, which is characteristic of a gas, becomes a disorderly motion of individual molecules in a "cage" formed by their closest neighbours. Although the motion of atoms in a gas differs considerably from that in a liquid, this difference is of rather a quantitative character and, in any case, the nature of the motion does not differ: in both cases the motion has a completely random character. From this it follows that under certain conditions the transition from a random motion at small densities to a random motion at large densities can occur gradually, without a jump at the condensation point. In other words, by changing the parameters p , V and T in a certain way one can get a continuous transition from the liquid state to the gaseous state and vice versa, without a discontinuous phase transition associated with the absorption or release of latent heat.

The possibility of a continuous transition between the liquid state and the gaseous state imposes an essential restriction upon the character of the liquid–gas phase equilibrium curve. Namely, a continuous transition between a liquid phase and a gaseous phase is possible only if the phase equilibrium curve $p(T)$ ends at a certain point C (fig. III.42), which is called the critical point (according to Mendeleyev – the point of absolute boiling).

Let p_c and T_c be the pressure and temperature at the critical point, called the critical pressure and critical temperature respectively. For all values of p and T lying below p_c and T_c the transition from the liquid into the gas and vice versa occurs with an intersection of the phase equilibrium curve. On the curve itself the two phases are in equilibrium with each other and are separated by a certain interface. Above the point C there is a uniform state of the substance in which there are no interfaces. This state is often called the

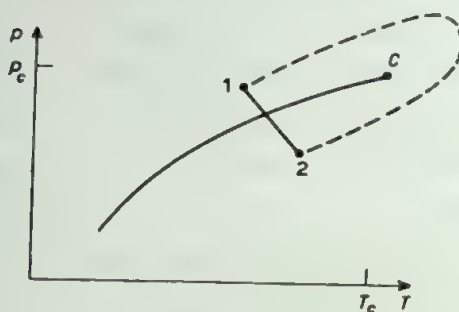


Fig. III.42

transcritical state. The uniform state of a substance can have a large or a small density, depending on the temperature and pressure. However, it makes no sense to call a substance above the critical point a liquid or a gas.

The transition from point 1 (liquid) to point 2 (gas) can be performed in the way shown by the solid line as well as in the way shown by the dotted line in fig. III.42. In the first way the phase equilibrium curve is intersected, so that the transition is accompanied by the release or absorption of latent heat. In the second way the transition takes place via the transcritical state and proceeds continuously, without the jump-like change in the character of the motion and without the release or absorption of latent heat. The possibility of a continuous transition from the liquid state into the gaseous state emphasizes the relative character of the terms "liquid" and "gas".

Strictly speaking, use can be made of the terms "liquid" and "gas" only when they exist simultaneously and are separated by an interface, i.e. when they are phases.

We shall find the conditions which determine the position of the critical point in the (p, T) -plane. Since it lies on the phase equilibrium curve, equilibrium conditions are fulfilled at it, in particular the conditions

$$\Delta p = 0, \quad \Delta T = 0,$$

where Δp and ΔT are the differences between the pressures and temperatures in the phases.

Near the critical point the difference between the two phases becomes small, and at the critical point itself it completely vanishes. In particular, the change in the density in the phase transition is very small and the densities

of the two phases are close to each other, in contrast to points lying large distances from the critical point.

If the difference between the densities of the phases is denoted by $\Delta\rho$, then one can always write the formal expansion

$$\Delta p = \left(\frac{\partial p}{\partial \rho}\right)_T \Delta\rho + \frac{1}{2} \left(\frac{\partial^2 p}{\partial \rho^2}\right)_T (\Delta\rho)^2 + \frac{1}{6} \left(\frac{\partial^3 p}{\partial \rho^3}\right)_T (\Delta\rho)^3 + \dots \quad (64.1)$$

Because of the phase equilibrium the sum of this series is equal to zero. Near the critical point the expansion is simplified. Sufficiently near the critical point $\Delta\rho$ can be assumed to be infinitesimal and in the expansion (64.1) one can drop higher terms of the expansion and write

$$\Delta p = \left(\frac{\partial p}{\partial \rho}\right)_T \Delta\rho = 0. \quad (64.2)$$

Since $\Delta\rho$ is an arbitrary infinitesimal quantity, it follows from (64.2) that

$$\left(\frac{\partial p}{\partial \rho}\right)_T = 0. \quad (64.3)$$

The derivative in (64.3) is taken at the critical point. Thus, at the critical point

$$\left(\frac{\partial p}{\partial \rho}\right)_T \sim \left(\frac{\partial p}{\partial V}\right)_T = 0. \quad (64.4)$$

If the quantity $(\partial p/\partial V)_T$ reduces to zero, then it is necessary for the stability of the substance that the following condition should simultaneously be fulfilled:

$$\left(\frac{\partial^2 p}{\partial V^2}\right)_T = 0. \quad (64.5)$$

Otherwise the fluctuations of the volume, as we have explained in § 57, would be infinitely large. The conditions (64.3) and (64.4) determine the position of the critical point. They are the same as the well-known conditions for the inflection point on the van der Waals curve.

Near the critical point the substance possesses a number of remarkable properties. The differences between the properties of the liquid and gaseous

phases progressively decrease and vanish at the critical point. It can be shown * that near the critical point the densities of the two phases depend on the temperature according to the law

$$\rho = \rho_c \pm [\text{const} \cdot (T - T_c)]^{\frac{1}{2}},$$

where the plus sign refers to the liquid and the minus sign to the gas. At the critical point the latent heat of transition reduces to zero, while the heat capacity C_p becomes infinite. The surface tension in the liquid-gas interface also reduces to zero at the critical point. The fact that the compressibility becomes infinite at the critical point leads to an essential decrease in the work of compression or expansion of a volume element in the system near the critical point.

Transitions from the crystalline state into an isotropic state have a different character. We understand an isotropic state to be the amorphous, liquid, or gaseous state. In this case a transition from a state of ordered motion into a state of chaotic motion takes place. To an ordered motion of atoms (or ions) in a crystal there corresponds the location of the atoms near the points of the crystal lattice and associated with this ordered distribution a definite crystal symmetry. As the temperature increases an increase occurs in the vibrational amplitudes of the atoms at the crystal lattice points and the number of violations of the regularity of the lattice increases, but the general ordered character of the motion and the symmetry of the lattice are preserved up to the melting point (or the sublimation point). A catastrophic disruption of the lattice occurs at the melting point, the symmetry vanishes, and the ordered motion is replaced by a chaotic motion.

In contrast to transitions between different isotropic phases, a continuous transition between a crystal and one of the isotropic phases is impossible. This impossibility is associated with a fundamental difference in the character of motion in these phases. This is also seen from symmetry considerations. It is impossible to transform, in a continuous way, an infinitely symmetric (isotropic) body into one with a definite finite symmetry. The same holds for phase transitions between different crystal modifications. Each modification possesses a definite ordered motion of atoms and a symmetry corresponding to this motion. In a phase transition the changes in the character of the motion and symmetry of the crystal occur discontinuously without fail. Owing to this a crystal \rightleftharpoons isotropic-phase equilibrium curve or a crystal \rightleftharpoons

* L.D. Landau and E.M. Lifshitz, *Course of theoretical physics*, Vol. 5: *Statistical physics* (Pergamon Press, London, 1958).

crystal equilibrium curve cannot have an end point and must go off to infinity. The transition from one phase into another is always associated with an intersection of the equilibrium curve and has the character of a jump*.

§65. Surface tension and surface pressure

Up to now we have considered the equilibrium of phases in contact, without taking into account the particular properties of the interface and their effect on the equilibrium. If, however, the phases which are in equilibrium possess a developed surface, a complete disregard of surface effects can introduce a fundamental error into calculations which are carried out. In this section we shall take into account the effect of the surface on the phase equilibrium.

Molecules which are distributed in a thin layer directly adjacent to the interface are in conditions which differ from those of molecules inside the main volume. They interact not only with molecules of their own phase but also with the contiguous layer of molecules of the other phase. Owing to this the structure and physical properties of the thin layer of substance, whose thickness is of the order of magnitude of the radius of the molecular interaction, turn out to be different from those of the main volume.

A detailed treatment of the properties of the surface layer would require a knowledge of the mechanism of the molecular interaction. Such a theory would be very complex. Therefore we have to simplify the problem, replacing the surface layer of finite thickness by an idealized infinitely thin interface which separates the two phases. Such an idealized infinitely thin surface layer we shall, for brevity, call the surface. The area of the surface is a new parameter characterizing the state of the system. For a given volume the system can have different values of the surface area Σ , a definite state of the system corresponding to each value of Σ .

A change in the surface of a system is accompanied by a gain or expenditure of energy. In order to form a new surface a particle from the volume must be brought to the surface, which requires that work be done. We denote by γ the generalized force corresponding to the parameter Σ .

If a change in the surface takes place at constant temperature, then the

* For more details about phase transitions see B.G. Levich, *Vvedenie v statisticheskuyu fiziku (Introduction to statistical physics)*, (Gostekhizdat, Moscow, 1954) §76, and for a particularly full exposition, L.D. Landau and E.M. Lifshitz, *Course of theoretical physics*, Vol. 5: *Statistical physics* (Pergamon Press, London, 1958).

work done in changing the surface ($dW = -\gamma d\Sigma$) is equal to the decrease in the free energy ($-dF_{\text{surf}}$), so that

$$dF_{\text{surf}} = \gamma d\Sigma . \quad (65.1)$$

The quantity γ , which represents the free energy per unit surface, is called the surface tension. The surface tension γ depends on the nature of the surface (in other words, on the nature of the phases forming it), as well as on the temperature. The value of γ does not depend, however, on the area of the surface. Hence one can write that

$$F_{\text{surf}} = \gamma \Sigma . \quad (65.2)$$

At constant temperature and volume there corresponds to an equilibrium state of an unclosed system the minimum value of the free energy. The phase contact area represents an example of such a system. Hence the phase contact area in an equilibrium state has the minimum possible value of the free energy. For a simultaneous change in the value of the surface and temperature in the system the change in the free energy of the surface has the form

$$dF_{\text{surf}} = -S_{\text{surf}} dT + \gamma d\Sigma . \quad (65.3)$$

From formula (65.3) it follows that the entropy of the surface is defined by the relation

$$S_{\text{surf}} = - \left(\frac{\partial F}{\partial T} \right)_{\Sigma} = \Sigma \frac{d\gamma}{dT} , \quad (65.4)$$

which allows one to express S_{surf} in terms of the surface tension.

In addition to the free energy of the surface one can write the expression for the energy of the surface E_{surf} :

$$E_{\text{surf}} = F_{\text{surf}} + TS_{\text{surf}} = \gamma \Sigma - T \frac{d\gamma}{dT} \Sigma = \left(\gamma - T \frac{d\gamma}{dT} \right) \Sigma . \quad (65.5)$$

This formula shows that it would be erroneous to define the surface tension as the energy per unit surface.

For the differential of the surface energy we can write

$$dE_{\text{surf}} = T dS_{\text{surf}} + \gamma d\Sigma . \quad (65.6)$$

From the above equality the relation

$$\left(\frac{\partial S_{\text{surf}}}{\partial \Sigma}\right)_{\Sigma_{\text{surf}}} = -\frac{\gamma}{T} \quad (65.7)$$

follows.

It should be noted that (65.4) cannot directly be substituted into (65.7). This would lead to the incorrect formula $d\gamma/dT = -\gamma/T$. The derivative in (65.7) is taken for a constant surface energy, but from (65.5) it follows that

$$\frac{d\gamma}{dT} \Sigma = \frac{\gamma \Sigma - E_{\text{surf}}}{T},$$

while from (65.4) it follows that

$$S_{\text{surf}} = -\frac{\gamma \Sigma - E_{\text{surf}}}{T}.$$

Differentiating the last expression for $E_{\text{surf}} = \text{const}$, we arrive again at formula (65.7).

The change in the surface energy can be resolved into the work done and the amount of heat absorbed:

$$dE_{\text{surf}} = TdS_{\text{surf}} + \gamma d\Sigma = dQ + dW.$$

Taking into account (65.4), it follows from this that the amount of heat which is absorbed as the area of the surface is reversibly increased by 1 cm^2 is equal to

$$Q = (T\Delta S)_{\Delta\Sigma=1} = -T \frac{d\gamma}{dT}. \quad (65.8)$$

The heat capacity of unit surface (for the constant value of the surface area $\Sigma = 1$) can be defined as

$$c_{\Sigma} = \left(\frac{\partial E}{\partial T}\right)_{\Sigma=1} = -T \frac{d^2\gamma}{dT^2}. \quad (65.9)$$

All quantities characterizing the thermodynamic properties of the surface are determined in terms of the surface tension and its derivatives with respect to the temperature.

The temperature dependence of the surface tension can at present be established theoretically only for quantum liquids (liquid helium II). It is impossible to separate the motion of the molecules of the surface layer from the motion of particles inside the liquid at high temperatures. Hence available attempts to calculate the surface tension without taking into account this connection are beneath criticism.

The surface tension changes the phase equilibrium condition. The condition of equilibrium in a system consisting of two phases and an interface for

$$dF = dF_1 + dF_2 + dF_{\text{surf}} = -p_1 dV_1 - p_2 dV_2 + \gamma d\Sigma = 0.$$

Since the volume of the entire system remains constant, $dV_2 = -dV_1$, so that

$$-(p_1 - p_2)dV_1 + \gamma d\Sigma = 0$$

or

$$p_1 = p_2 + \gamma \frac{d\Sigma}{dV_1}. \quad (65.10)$$

The quantity $d\Sigma/dV_1$ represents the curvature of the interface

$$\frac{d\Sigma}{dV_1} = \frac{1}{r_1} + \frac{1}{r_2},$$

where r_1 and r_2 are the major radii of curvature. In the case of a spherical surface

$$\frac{d\Sigma}{dV_1} = \frac{d(4\pi r^2)}{d(\frac{4}{3}\pi r^3)} = \frac{2}{r}.$$

In this case the radius vector is assumed to be positive if it is directed towards the first phase. Thus, we have finally

$$p_1 = p_2 + \gamma \left(\frac{1}{r_1} + \frac{1}{r_2} \right). \quad (65.11)$$

The quantity $\gamma(r_1^{-1} + r_2^{-1})$ is called the Laplace pressure, and formula (65.11) is called the Laplace formula. The Laplace formula shows that the pressure

in the first phase is balanced by the sum of the pressure in the second phase and the Laplace pressure. In the particular case of a plane interface the Laplace pressure reduces to zero, since for a plane surface $r_1 \rightarrow \infty$, $r_2 \rightarrow \infty$.

The presence of surface tension changes not only the condition of mechanical equilibrium but also the phase equilibrium condition for particle exchange.

That is, although particles cannot be detained in the interface in the case of phase equilibrium and can go from one phase into the other without hindrance, condition (61.4) will now be fulfilled for a somewhat changed value of the pressure (in comparison with the pressure at the point of phase transition at the same temperature for a plane interface).

Let us consider the case of phase equilibrium between the drops of a liquid and its vapour. We shall assume the drops to be spheres of radius r .

The equilibrium condition for the particle exchange will have the form

$$\mu_1(p'', T) = \mu_2(p', T), \quad (65.12)$$

where p'' is the pressure of the saturated vapour, p' is the corresponding pressure in the liquid drop, and T is the transformation temperature. We assume the latter to be equal to the transformation temperature in the case of a plane surface. The same condition at a plane surface has the form

$$\mu_1(p, T) = \mu_2(p, T). \quad (65.13)$$

Subtracting (65.13) from (65.12), we find

$$\mu_1(p'', T) - \mu_1(p, T) = \mu_2(p', T) - \mu_2(p, T). \quad (65.14)$$

Since the compressibility of the liquid is very small, the difference between the pressures p' and p is also small. In view of this one can write that

$$\begin{aligned} \mu_2(p', T) - \mu_2(p, T) &= \mu_2(p + \Delta p, T) - \mu_2(p, T) = \\ &= \frac{\partial \mu_2}{\partial p} \Delta p = v_2 \Delta p, \end{aligned} \quad (65.15)$$

where v_2 is the volume per particle in the liquid phase, and Δp is the difference between the pressures in the liquid for a spherical surface and plane surface. This difference is obviously equal to

$$\Delta p = \frac{2\gamma}{r}.$$

The change in the pressure of the saturated vapour above a spherical drop in comparison with a plane surface is in general not small, and the left-hand side of (65.14) cannot be expanded in a series. We shall rewrite it, making use of the general formula for the chemical potential (60.5) of an ideal gas. Assuming the vapour to be an ideal gas, we find

$$kT \ln \frac{p''}{p} = v_2 \frac{2\gamma}{r}. \quad (65.16)$$

Formula (65.16) relates the pressure of the saturated vapour above the drop to its radius. Writing it in the form

$$p'' = p \exp \left[\frac{2v_2\gamma}{rkT} \right], \quad (65.17)$$

we see that p'' rapidly increases with decreasing radius of the drop, and that for small drops it can become considerable.

For example, for water drops with $\gamma \approx 80 \text{ erg/cm}^2$, $r = 10^{-6} \text{ cm}$ and $T \approx 300 \text{ K}$ the pressure $p'' \approx 1.1p$, i.e. the pressure of the vapour above the drop exceeds that above a plane surface by 10%.

Formula (65.17) shows that a system consisting of a set of drops of different size is in a state of unstable equilibrium. Small drops possessing an excess energy which is associated with the surface tension will evaporate, while the vapour will condense on big drops. This process, which is called the distillation of drops, will go on until the entire liquid transforms into drops of the largest size.

Another important phenomenon, which is associated with the change in the pressure of the vapour above a curved surface, is observed in capillary tubes wetted by a liquid. In such capillary tubes the surface of the meniscus will be concave, so that one has to write a minus sign in formula (65.17). The pressure of the saturated vapour above a concave surface turns out to be smaller than that above a plane surface. Owing to this, condensation of the vapour takes place in fine capillary tubes before it does above a plane surface at the same temperature. This phenomenon is called capillary condensation.

Formula (65.17), which has been derived for equilibrium in a liquid-vapour system, is qualitatively valid also in other cases of phase equilibrium, for example, a crystal-vapour system.

§66. Gas adsorption

One of the most important effects associated with the particular properties of the phase contact area is adsorption. By adsorption is meant the accumulation of a substance on the surface of a solid or liquid phase. As a rule the adsorbed substance is distributed over the phase contact area and does not penetrate at all into the condensed phase. The phenomenon of adsorption is observed for many different combinations of phases. In practice one most often has to deal with the adsorption of gases on the surface of a solid body. In this section we shall confine ourselves to the treatment of this case of adsorption.

For the present, the nature of the forces which bind molecules adsorbed on a surface with the molecules of the solid or liquid backing (called the adsorbent) cannot be considered as well established. In a number of cases they have the character of van der Waals forces, in other cases a stronger bond is established between the adsorbed molecules and the molecules of the adsorbent, corresponding to the formation of a particular chemical combination. Intermediate cases are also possible. As a rule, adsorbed molecules are distributed on the surface of the adsorbent in the form of a monomolecular layer. Adsorbed molecules most often possess no mobility, and in the process of adsorption they are fixed to quite definite points on the surface of the crystal. We shall call such points on the surface of a crystal at which the adsorption of molecules takes place sites. The sites can be the edges of the faces of a crystal or any other outstanding points on the surface.

We denote the number of sites per cm^2 of the surface of the crystal by N_L . We assume that all sites on the surface are equivalent, so that the adsorbed molecule is bound to the surface at each of them to the same degree.

We write the partition function for a system of adsorbed particles, assuming that their density N_A (the number of particles per cm^2 of the surface) is small and that the interaction between adsorbed molecules can be disregarded. Every one of the adsorbed particles possesses a potential energy $(-u)$ which is a measure of the work done in removing the particle from the surface.

If it is assumed that the adsorbed molecules are oscillating about an equilibrium position with a frequency ν , then the partition function for each molecule can be written in the form

$$z_A = e^{u/kT} z_{\text{oscil}},$$

where z_{oscil} is the partition function of the oscillator.

The partition function for the entire set of adsorbed particles can be written in the form

$$Z_A = \frac{N_L!}{N_A!(N_L - N_A)!} (z_A)^{N_A}, \quad (66.1)$$

where the factor $\frac{N_L!}{N_A!(N_L - N_A)!}$ represents the statistical weight of the state.

Indeed, to a given energy of the system there correspond a number of states which differ in the distribution of the particles over sites. The number of ways in which N_A particles can be distributed over N_L sites is equal to $N_L!/[N_A!(N_L - N_A)!]$. It is this number which gives the statistical weight of a state of a given energy of the system of particles.

By means of (66.1) one can find the free energy and the chemical potential of the system. We have for the free energy

$$\begin{aligned} F_A &= -kT \ln Z_A = -kT \ln \left\{ \frac{N_L!}{N_A!(N_L - N_A)!} (z_A)^{N_A} \right\} \approx \\ &\approx -N_L kT \ln \frac{N_L}{e} + kT N_A \ln \frac{N_A}{e} + \\ &+ (N_L - N_A) kT \ln \frac{N_L - N_A}{e} - N_A kT \ln z_A. \end{aligned} \quad (66.2)$$

Correspondingly, the chemical potential of adsorbed particles is equal to [see (60.3')]

$$\begin{aligned} \mu_A &= \frac{\partial F_A}{\partial N_A} = kT \ln \frac{N_A}{e} - kT \ln \frac{N_L - N_A}{e} - kT \ln z_A = \\ &= kT \ln \frac{N_A}{N_L - N_A} - kT \ln z_A. \end{aligned} \quad (66.3)$$

By means of formula (66.3) one can consider the equilibrium between an adsorbed substance and the gas. Equating μ_A and the chemical potential of the gas given by formula (60.5), we obtain

$$kT \ln \frac{N_A}{N_L - N_A} + f(T) = kT \ln p + \varphi(T),$$

whence we find

$$\frac{N_A}{N_L - N_A} = \frac{p}{p_0(T)}, \quad (66.4)$$

where p_0 depends only on the temperature. Solving (66.4) for N_A , we obtain

$$N_A = N_L \frac{p}{p + p_0(T)}. \quad (66.5)$$

Formula (66.5) represents the adsorption isotherm: it determines the number of adsorbed molecules as a function of the pressure of the gas above the surface of a solid body at a given temperature. It is obvious that at low pressure, when $p \ll p_0$, this number is proportional to the pressure of the gas and to the number of sites:

$$N_A \approx N_L \frac{p}{p_0(T)}.$$

In this case the degree of occupation of the sites is small. At a large pressure $p \gg p_0$ the phenomenon of saturation occurs, the number of adsorbed molecules ceases to depend on the pressure and becomes a constant equal to the number of sites:

$$N_A \approx N_L.$$

The adsorption isotherms of formula (66.5) are shown in fig. 111.43.

At high temperatures a mobility of the adsorbed molecules across the surface arises. In the limit the adsorbed molecules can move across the surface of a solid body like the molecules of a "two-dimensional" gas. Then the form of the partition function z_A of the adsorbed molecule changes, but the form of the adsorption isotherm remains as before.

It should be noted that the simplest mechanism of adsorption considered here, for which all sites on the surface are characterized by one and the same binding energy ($-u$) and which leads to the isotherm (66.5), is seldom encountered. Usually there are different sites on the surface of a solid body with different values of the binding energy, i.e. the surface of the adsorbent is non-uniform. Moreover, when sites are densely occupied the interaction between molecules in the adsorbed layer becomes important and affects the shape of the isotherm. As a result of the superposition of these factors the

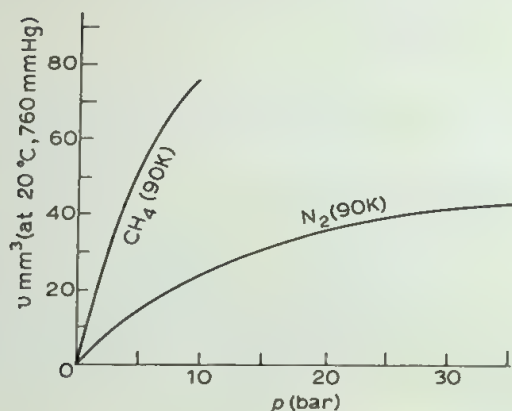


Fig. III.43

form of the isotherm can change and be very different from that shown in fig. III.43. By determining the form of the isotherm from experimental data and considering it as a result of the superposition of the isotherms (66.5) with different $p_0(T)$, one can obtain information about the non-uniformity of the surface.

As to gas adsorption on the surface of a liquid, the mechanism of adsorption here does not differ from that on the surface of a solid body at high temperatures; all points on the surface of a liquid are equivalent and the adsorbed particles are mobile on the surface for a relatively sparse population. The adsorption isotherm has the form (66.5). When the surface is very densely populated the interaction between the adsorbed molecules, which distorts the form of the isotherms, begins to play an important role.

§67. Chemical equilibrium in the gas phase

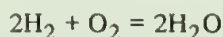
As the second example of a system with a variable number of particles we shall consider equilibrium in a system in which a chemical reaction occurs.

We assume for concreteness that in the course of the chemical reaction atoms A and B combine to form a molecule AB. The molecule AB in its turn dissociates into individual atoms A and B. Both processes, that of the combination and that of the dissociation, have a certain rate, by which is meant the number of events occurring per unit time. The rate of the direct process

$A+B \rightarrow AB$ is in general not equal to that of the reverse process $AB \rightarrow A+B$. Hence the chemical reaction proceeds mainly in one direction. However, after the lapse of a certain time, when reagents arising in the course of the more rapid reaction have accumulated while the amount of reagents vanishing in the course of the reaction have decreased the rate of the rapid reaction will decrease whereas that of the slow reaction will increase. As a result an equilibrium state will be established in the system. The number of molecules AB produced and disintegrated will be the same. In this case one speaks of the equilibrium reaction $A+B \rightleftharpoons AB$. It is more convenient to write the reaction in the form of an equality. If the reaction takes place between several substances, then each chemical reaction can be written in the form

$$\nu_i g_i = 0, \quad (67.1)$$

where the g_i are the chemical symbols of the reacting substances, and the ν_i are the number of reacting moles of the corresponding substances. We choose to write reactions in such a way that the coefficients ν_i for substances which are used up in the course of the reaction have a negative sign, while those for substances produced have a positive sign. For example, the reaction of the formation of water vapour from the mixture



must from this point of view be written in the form

$$2H_2O - 2H_2 - O_2 = 0,$$

so that

$$\nu_{H_2O} = 2, \quad \nu_{H_2} = -2, \quad \text{and} \quad \nu_{O_2} = -1.$$

Let us write the conditions of chemical equilibrium in an arbitrary system consisting of initial substances and reaction products. The set of molecules of the initial substance and the set of molecules of the reaction products can be considered as certain quasi-closed systems which are in a reservoir and are weakly interacting with each other. The latter condition is fulfilled if the number of atoms reacting per unit time is small in comparison with the total number of molecules in the system, which is always the case in a macroscopic system of substances for an equilibrium reaction.

One usually investigates the equilibrium states of a reacting system at a given temperature and pressure. The equilibrium condition is the requirement

$$G(p, T, N_i) \rightarrow \min ,$$

where N_i is the number of particles of a given kind. For constant given values of the temperature and pressure in the entire system the condition of the minimum can be rewritten in the form

$$dG = \left(\frac{\partial G}{\partial N_1} \right)_{p,T} dN_1 + \left(\frac{\partial G}{\partial N_2} \right)_{p,T} dN_2 + \dots = \sum \mu_i dN_i = 0$$

and, noting that the change in the number of particles of a given kind can be written in the form

$$dN_i = \nu_i dN ,$$

we obtain the equilibrium condition in a system in the presence of chemical reactions:

$$\sum \nu_i \mu_i = 0 . \quad (67.2)$$

When one molecule of the first subsystem transforms into one molecule of the second subsystem (the case which we have considered before) the coefficients ν_i are obviously equal to $\nu_1 = 1$, $\nu_2 = -1$. In this case formula (67.2) turns out to be identical with (61.4).

We see that chemical equilibrium is determined by the equality of the chemical potentials.

§68. The law of mass action

In order to apply the conditions (67.2) to definite cases of chemical equilibrium it is necessary to know the explicit form of the chemical potentials. The latter are known mainly for gases. Therefore the theory which follows will refer to chemical equilibrium in a mixture of gases. We have, in §60, calculated the chemical potential of a gas. In a mixture of ideal gases each of the gases behaves as though it alone occupied the entire volume of the container and has a chemical potential $\mu_i = N_i p_i / N$, where N_i is the number of particles of the i th gas and N is the total number of atoms of all kinds in the container. We write μ_i in the general form:

$$\mu_i = kT \ln p_i + \chi_i(T), \quad (68.1)$$

where

$$\begin{aligned} \chi(T) &= -\frac{5}{2}kT \ln kT - kTj_1, \\ j_1 &= \ln \left[\left(\frac{2\pi m}{h^2} \right)^{\frac{3}{2}} \right] \end{aligned} \quad (68.2)$$

for a monatomic gas,

$$\chi(T) = -\frac{7}{2}kT \ln kT - kTj_2 + kT \ln (1 - e^{-h\nu/kT}) + \epsilon_0, \quad (68.3)$$

$$j_2 = \ln \left[\left(\frac{2\pi m}{h^2} \right)^{\frac{3}{2}} \frac{8\pi^2 I}{\gamma h^2} \right]$$

for a diatomic gas at very high temperatures when vibrations are excited, and

$$\chi(T) = -\frac{7}{2}kT \ln kT - kTj_2 + \epsilon_0$$

for a diatomic gas at not very high temperatures, when vibrations are not excited.

Let us consider a reaction of the type

$$\nu_1 g_1 + \nu_2 g_2 - \nu_3 g_3 = 0.$$

The chemical equilibrium condition reads

$$\nu_1 \mu_1 + \nu_2 \mu_2 - \nu_3 \mu_3 = 0 \quad (68.4)$$

or

$$\nu_1 kT \ln p_1 + \nu_2 kT \ln p_2 - \nu_3 kT \ln p_3 = \nu_3 \chi_3 - \nu_1 \chi_1 - \nu_2 \chi_2. \quad (68.5)$$

Thus,

$$\frac{p_1^{\nu_1} p_2^{\nu_2}}{p_3^{\nu_3}} = K(T), \quad (68.6)$$

where

$$\ln K(T) = \frac{\nu_3 X_3 - \nu_1 X_1 - \nu_2 X_2}{kT} . \quad (68.7)$$

The quantity $K(T)$ depends only on the temperature and nature of the reacting molecules but not on the initial pressures or the amount of the reacting gases.

Formula (68.6) is called the law of mass action. The law of mass action shows that, irrespective of the initial composition of the reacting gas mixture, in the course of time there is established in it an equilibrium state for which the partial pressures have quite definite values which are related by formula (68.6). They do not depend on any parameters except the temperature, the difference between the zero point energies, and the chemical constants of the reacting gases. In the case where the reaction involves not three but a larger number of gases the mass action law must be written in the form

$$\prod \frac{(p_{i'})^{\nu_i'}}{(p_i)^{\nu_i}} = K(T) , \quad (68.8)$$

where the product is taken over all the gases taking part in the reaction, and the primes refer to reaction products.

The law of mass action was first discovered experimentally by N.N.Beketov, and it was introduced theoretically, on the basis of statistical considerations, by Guldberg and Waage.

The expression (68.8) has a clear statistical meaning: in order that the initial substances may react it is necessary that their molecules should simultaneously be in a very small volume v , whose size is of the order of magnitude of the diameter of the molecules. Since the gases are assumed to be ideal and the motion of the molecules is independent of one another, the probability that the molecules of the initial substances will simultaneously be found in a given volume is proportional to the numbers of these molecules in the gas. These numbers are in their turn proportional to the corresponding partial pressures. Thus, the probability of the direct reaction w_1 is proportional to $p_1^{\nu_1} p_2^{\nu_2} \dots$

$$w_1 = ap_1^{\nu_1} p_2^{\nu_2} \dots .$$

The same reasoning can also be applied to the reverse reaction. The probability of the reverse reaction w_2 is equal to

$$w_2 = b(p_1')^{\nu_1'} (p_2')^{\nu_2'}.$$

In an equilibrium state the rate of the direct reaction is equal to that of the reverse reaction. For this the probability of the direct process must be equal to that of the reverse process. Equating w_1 and w_2 and denoting by K the ratio of the factors of proportionality a/b , we arrive at formula (68.8).

The mass action law represents the basic law of chemical equilibrium. It can be derived in a purely thermodynamic way, but then the value of the constant $K(T)$ remains indefinite and must be found experimentally. By means of the statistical expressions for the μ_i which have been presented above, the constant K can be calculated theoretically. We shall give an example of such a calculation somewhat later.

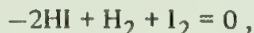
The law of mass action is most often expressed not in terms of partial pressures but in terms of the so-called molar fractions:

$$c_i = p_i/p.$$

Substituting c_i into (68.6), we obtain

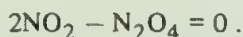
$$\frac{c_1^{\nu_1} c_2^{\nu_2}}{c_3^{\nu_3}} = p^{-\sum \nu_i} K(T). \quad (68.9)$$

From formula (68.9) it follows that, if the reaction proceeds without a change in the number of moles, so that $\nu_3 = \nu_1 + \nu_2$, the equilibrium does not depend on the total pressure p in the system. As an example of such a reaction we point to the reaction of the dissociation of hydrogen iodide



for which $\nu_1 = 1$, $\nu_2 = 1$, $\nu_3 = -2$.

If the reaction proceeds with a change in the number of moles, so that $\nu_3 \neq \nu_1 + \nu_2$, then the change in the total pressure shifts the equilibrium. This means that the ratio between the molar fractions of the initial substance and the reaction product changes as the total pressure changes. Let, for example, the dissociation of the N_2O_4 molecule into two NO_2 molecules takes place. We write the reaction in the form



The coefficients of the reaction are $\nu_{\text{NO}_2} = 2$, $\nu_{\text{N}_2\text{O}_4} = -1$, so that the reac-

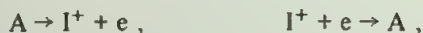
tion proceeds with an increase in the number of moles. The law of mass action reads:

$$\frac{c_{\text{NO}_2}^2}{c_{\text{N}_2\text{O}_4}} = p^{-1} K(T) .$$

As the total pressure decreases the number of NO_2 molecules increases, i.e. the percentage of dissociated N_2O_4 molecules in the equilibrium mixture decreases. Thus, if the reaction proceeds with an increase in the number of moles, $|\nu_3| > \nu_1 + \nu_2$, then a decrease in the total pressure favours the reaction, whereas an increase in the pressure hampers it. In the case of reactions proceeding with a decrease in the number of moles, $|\nu_3| < \nu_1 + \nu_2$, the change in the total pressure acts in the reverse direction.

§69. Thermal dissociation of atoms

We have mentioned earlier the thermal dissociation of atoms which occurs at very high temperatures. When the temperature reaches such high values that the thermal energy kT becomes comparable with the energy required to tear an electron away from an atom (the ionization energy), then thermal ionization of atoms takes place. Atoms dissociate into positively charged ions and electrons which form the corresponding ideal gases. In addition to the process of ionization the reverse process also takes place: recombination, in the course of which an ion and an electron combine to form a neutral atom. Thus, at very high temperatures two reactions take place in the substance:



where A denotes the atom and I^+ denotes the ion (for simplicity we confine ourselves to the case of single ionization). If constant conditions of temperature and pressure are maintained in the system, then an equilibrium state will be established in which the number of dissociations is equal to the number of recombinations. A system in which an equilibrium reaction



takes place in principle does not differ at all from a system in which an equilibrium chemical reaction takes place. For such a system the law of mass action can be written in the form

$$\frac{c_{I^+} c_e}{c_A} = p^{(\nu_A - \nu_{I^+} - \nu_e)} K(T) = \frac{K(T)}{p}, \quad (69.1)$$

where c_e , c_{I^+} , c_A are the molar fractions of the electron gas, ion gas and atomic gas respectively. The constant $K(T)$ is given by

$$\ln K(T) = \frac{5}{2} \ln kT + j_e - \frac{\Delta\epsilon_0}{kT}. \quad (69.2)$$

In formula (69.2) all quantities referring to the ion and the atom have been cancelled, since the difference between the mass of the ion and the mass of the atom can be neglected. The chemical constant of the electron gas is equal to

$$j_e = \ln \left[\left(\frac{2\pi m_e}{h^2} \right)^{\frac{3}{2}} \right].$$

The quantity $\Delta\epsilon_0$ represents the ionization energy of the atom. Thus,

$$\frac{c_{I^+} c_e}{c_A} = \frac{1}{p} \exp \left(\frac{5}{2} \ln kT + j_e - \frac{\Delta\epsilon_0}{kT} \right) = \frac{K(T)}{p}. \quad (69.3)$$

Instead of molar fractions it is more convenient to introduce another, more obvious quantity, which is called the degree of dissociation. Let an α th part of atoms undergo dissociation, so that $N(1+\alpha)$ particles arise out of N atoms. The quantity α characterizing the fraction of ionized atoms is called the degree of dissociation. Obviously, we have

$$c_{I^+} = \frac{\alpha}{1+\alpha}, \quad c_e = \frac{\alpha}{1+\alpha}, \quad c_A = \frac{1-\alpha}{1+\alpha}.$$

Substituting molar fractions expressed in terms of the degree of dissociation into (69.3), we find

$$\frac{\alpha^2}{1-\alpha^2} = \frac{K}{p},$$

whence

$$\alpha = \left[1 + \frac{p}{K} \right]^{-\frac{1}{2}} = \left[1 + \frac{p}{(kT)^{\frac{5}{2}}} \left(\frac{h^2}{2\pi m} \right)^{\frac{3}{2}} \exp \frac{\Delta \epsilon_0}{kT} \right]^{-\frac{1}{2}}. \quad (69.4)$$

From this formula it is seen that the degree of dissociation increases rapidly with temperature. Estimates of the numerical values of the quantities contained in the coefficient of the exponential factor in the denominator show that this factor is very small.

Hence, provided $\Delta \epsilon_0/kT$ is not very large, the entire expression under the square root, and also α , is of the order of unity. This means that for values of kT which are comparable with the ionization energy the gas turns out to be practically completely ionized:

$$\alpha \approx 1; \quad c_{I^+} \approx \frac{1}{2}; \quad c_e \approx \frac{1}{2}; \quad c_A \approx 0.$$

The degree of ionization also increases with decreasing total pressure. This is in complete agreement with what was said at the end of §68 for reactions proceeding with an increase in the number of moles.

A well-known application of formula (69.4) is in the elucidation of a peculiar feature of the spectrum of the solar atmosphere which at first sight seems to be very strange. A method of investigating spectra which originate from different layers of the solar atmosphere (chromosphere) has been developed in astrophysics. Investigations have shown that in the deeper layers of the atmosphere, where the temperature is higher, the degree of dissociation of the vapour of calcium is lower than that in outer, cooler layers. In the case of calcium the ionization potential amounts to 6 eV. The degree of ionization α at 6000 K and a pressure $p = 1$ atm amounts to only 8%, whereas at the same temperature and a pressure of 10^{-2} atm it reaches 65%. The explanation lies in the fact that, owing to the effect of the coefficient of the exponential term containing p in formula (69.4), the increase in the degree of dissociation with decreasing pressure is more rapid than its decrease with decreasing temperature in going from the deeper to the outer layers of the solar atmosphere.

Statistical Distributions in Quantum Statistics and Some of Their Applications

§70. The identity of elementary particles and the calculation of the partition function

We have already pointed out more than once that classical concepts turn out to be inadequate for the study of the motion of atomic systems and that they must be replaced by the concepts of quantum theory. In ch. I we presented the minimum amount of information from quantum theory which was necessary for the subsequent exposition. However, for a more profound analysis of those changes which are introduced into statistical physics by quantum theory it is necessary to dwell on certain important results of quantum theory.

As we have already stressed, the following two propositions of quantum mechanics are of basic importance for statistical physics:

- (1) the existence of the discrete states of a system,
- (2) the principle of identity of elementary particles.

We have from the very beginning taken into account the discreteness of quantum states. We have established when it is necessary to take into account the discrete character of energy levels, and when they can be considered approximately as distributed continuously. Also, we have shown the effect of the discreteness of the energy spectrum on the behaviour of statistical

systems. However, up to now we have not taken the identity of particles into account in a systematic way. True, we have considered as one state those states differing from each other only by a permutation of the particles. For this we have divided the phase space by the number of possible permutations of the particles. This division represented the simplest attempt of taking into account the identity of particles. As a matter of fact, the division by $N!$ was carried out even before the appearance of quantum theory. Otherwise, as we have already pointed out in §37, incorrect expressions for thermodynamic functions were obtained from statistics. Considerations based on the principle of identity of elementary particles justified to a certain degree the division of the partition function by $N!$. However, the inconsistency of this operation is obvious. Indeed, we have assumed in the beginning that all particles are different from each other, so that one can in principle number them, ascribing a definite number or label to each particle. Proceeding from this point of view, we have calculated the possible states of a system consisting of N independent particles, integrating with respect to the coordinates and momenta of the first, second and so on particles. Thereupon, in contradiction with the initial assumption of the possibility of numbering the particles, we have proclaimed that the part of the states differing from each other only by a permutation of the particles to be completely identical and have required that all of them should be taken into account as one.

Experiment and theory show, however, that the identity of atomic particles has a much more profound character. The complete identity of atomic particles leads to the fact that the first of the operations which we have carried out — the numbering of particles — loses any physical meaning. It makes no sense to call one of the particles the first, another the second and so on and to integrate them over states, since there are no physical differences between the first, second and so on particles. If one calls the particle which is at the initial instant in a definite state the first particle, then at a subsequent instant it would already be impossible to assert that it is just the first particle which is in this state, because it would be impossible to distinguish the first particle from a "not first" particle. Hence it is necessary from the very beginning to renounce any attempt to distinguish between individual atomic particles, i.e. to renounce the property of a system of atomic particles which we have adopted *.

We shall now see what changes in the statistical distribution result from taking account of the complete identity of atomic particles. All further discussion will refer only to a monatomic ideal gas.

* Exceptions are the so-called systems of localized particles which are separated from each other by impenetrable barriers. We shall not consider such systems.

§ 71. A second method of deriving the statistical distribution

In order to derive the statistical distribution in a gas, taking into account the principle of the identity of elementary particles, we shall have recourse to a special method which is characteristic of the versatility of statistical methods. To make the difference between the classical and quantum considerations particularly striking, we shall first introduce the classical distribution (the Maxwell distribution) by means of this method.

Assume that the molecules in a gas can be in individual quantum states with energies of translational motion $\epsilon_1, \epsilon_2, \epsilon_3, \dots$ (for the convenience of the treatment we shall for the present assume that the energy levels are discrete). In the gas there is a certain distribution of particles over states, such that in the first state there are n_1 particles, in the second state there are n_2 particles, and so on. From the classical point of view we have to describe the states of the gas in the following way:

particles No. 1, 2, 3, ..., n_1 are in the state with energy ϵ_1 ,

particles No. $n_1 + 1, \dots, n_1 + n_2$ are in the state with energy ϵ_2 ,

particles No. $n_1 + n_2 + 1, \dots, n_1 + n_2 + n_3$ are in the state with energy ϵ_3 and so on.

Further, we choose as a quasi-closed subsystem all particles which are in a certain arbitrarily chosen quantum state with energy ϵ_k . All the remaining gas particles, which are in other energy states, then form a reservoir.

The choice of a group of particles which are in a given state as a subsystem is in complete agreement with those requirements which must be satisfied by a quasi-closed subsystem (§ 13). Indeed, as a result of collisions the particles possessing the energy ϵ_k go over into other states. Conversely, owing to the same mechanism other molecules, which earlier had an energy differing from ϵ_k and which, consequently, belonged to the reservoir, can go into the state with energy ϵ_k . If the number of molecules which are coming in or going out of the subsystem per unit time is small in comparison with the number of particles in it, then it can be assumed that the interaction between the subsystem and the reservoir is weak. This condition will be fulfilled if the collisions between the particles which cause the corresponding transitions occur sufficiently seldom, i.e. when the gas is rarefied. Since the interaction of the subsystem with the reservoir consists in the passage of particles from the subsystem to the reservoir and vice versa, the subsystem which we have chosen represents an example of a subsystem with a variable number of particles. It differs from the general case of a subsystem with a variable number of particles and a variable energy by the fact that the energy of each particle in the subsystem is fixed. However, the energy of the subsystem,

which is made up of the energies of all the particles contained in it, of course, also changes as the number of particles in it changes.

In order to avoid misunderstanding we stress that we are now speaking not of the states of an actually existing system but of the states of a conditionally introduced subsystem. Our subsystem is not a single system but a set of particles with a definite energy which are at different loci of the gas and are not bound to each other.

In the process of variation of a state of a real system the number of particles getting into the state with the given energy ϵ_k changes. In this sense the state of our subsystem varies. The energy of the subsystem is equal to

$$\epsilon = \epsilon_k n_k, \quad (71.1)$$

where n_k is the number of particles in the subsystem. The quantity ϵ varies together with n_k .

In order to characterize completely the state of the subsystem we have chosen, it is necessary to know the mean number of particles in it, i.e. the mean number of particles which are at the energy level we have chosen. For this calculation use can be made of the general formula (59.11) which gives the mean number of particles in a subsystem with a variable number of particles. In our special case this formula can be essentially simplified. We do not have to introduce the double summation over the possible values of the energy and the number of particles, because in our system the value of the energy is unambiguously determined by the number of particles contained in it according to formula (71.1). In summing over the possible values of the number of particles in the subsystem we automatically carry out the summation over the possible values of its energy. As a matter of fact, it is with this simplification that our choice of the subsystem has been associated. Thus, the mean number of particles \bar{n}_k in the subsystem is expressed by the formula

$$\bar{n}_k = kT \frac{\partial}{\partial \mu} \ln \sum_{n_k} \left(\exp \frac{\mu - \epsilon_k}{kT} \right)^{n_k} \Omega(n_k), \quad (71.2)$$

where instead of ϵ we have substituted its expression according to formula (71.1).

The summation in (71.2) is carried out over all possible values of the number of particles n_k in the subsystem. In order to actually perform the summation in formula (71.2) it is necessary to know the explicit expression for the statistical weight (the number of states) $\Omega(n_k)$ of the state of the system when it contains n_k particles. In quasi-classical statistics, where all

particles can be numbered, the states of our system will always be degenerate, provided the system contains more than one particle of a given kind. Indeed, if the system contains N identical particles, then it can be in states which differ from each other by a permutation of the particles. For example, let there be in our gas two molecules, No. 1 and No. 2, with energy ϵ_k . In the first state molecule No. 1 is at point 1 while molecule No. 2 is at point 2. In the second state the positions of the molecules are exchanged. The energy of the subsystem in the two states is the same and equal to $2\epsilon_k$. Thus, there are two states of the system with an energy $2\epsilon_k$ or, in other words, the states of the system are two-fold degenerate. In the general case the states of a system containing n_k particles are $n_k!$ -fold degenerate.

If we do not wish to consider states differing only by a permutation of the particles as different states (which would undoubtedly lead us to incorrect expressions for the thermodynamic functions), then for Ω use should be made of the general formula (1.26) and the total volume of the phase space must be divided by the number of possible permutations of molecules $n_k!$ and the size h^3 of a cell.

The volume of phase space corresponding to one quantum state with an energy ϵ_k is obviously equal to h^3 . Hence for $\Omega(n_k)$ we finally obtain

$$\Omega(n_k) = \frac{1}{n_k!}. \quad (71.3)$$

Substituting the expression (71.3) into formula (71.2), we find

$$\bar{n}_k = kT \frac{\partial}{\partial \mu} \ln \sum_{n_k=0}^N \frac{1}{n_k!} \left(\exp \frac{\mu - \epsilon_k}{\theta} \right)^{n_k}. \quad (71.4)$$

The number of particles n_k contained in the subsystem can vary from zero to the total number of particles in the gas N . However, the probability that all particles of the gas be found simultaneously in one energy state is extremely low. Hence for n_k which are close to N the terms of the sum (71.4) are so small that the sum rapidly converges. Therefore we shall not commit an error if we replace the upper limit N in the sum by infinity. This corresponds to the addition of infinitesimal terms to the sum. For such a substitution the sum (71.4) goes over into a simple series:

$$\sum_{n_k=0}^{\infty} \frac{1}{n_k!} \left(\exp \frac{\mu - \epsilon_k}{\theta} \right)^{n_k} = \sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x, \quad (71.5)$$

where we have provisionally denoted $\exp(\mu - \epsilon_k)/\theta$ by x . Thus,

$$\bar{n}_k = kT \frac{\partial}{\partial \mu} \ln e^x = kT \frac{\partial}{\partial \mu} \exp \frac{\mu - \epsilon_k}{kT} = \exp \frac{\mu - \epsilon_k}{kT}. \quad (71.6)$$

In particular, if the states of the molecule and its energy vary continuously, which is always valid in classical statistics, then instead of a given energy level ϵ_k one has to consider states with an energy lying between ϵ and $\epsilon + \delta\epsilon$. Then instead of the number of particles n_k in a given quantum state one needs an expression for the mean number of particles with an energy between ϵ and $\epsilon + \delta\epsilon$, which we shall denote by dn . Obviously,

$$dn = \bar{n} \frac{d\gamma}{h^3} = \left(\exp \frac{\mu - \epsilon}{kT} \right) \frac{d\gamma}{h^3}, \quad (71.7)$$

where $d\gamma$ is the volume of phase space corresponding to an energy between ϵ and $\epsilon + \delta\epsilon$, and $d\gamma/h^3$ is the number of states with this energy. Formula (71.7) is the same as the Maxwell distribution in the form it had in formula (60.10).

In later sections we shall make use of the statistical distribution (71.2) to obtain the quantum laws of the distribution of molecules in an ideal gas.

In conclusion we note that the method of deriving the Maxwell-Boltzmann distribution described here is often called the method of cells in phase space.

§ 72. Quantum distributions for an ideal gas

As we have just stressed, it follows from the principle of identity of particles that one cannot distinguish between individual microscopic particles; electrons, photons, protons and other elementary particles or atoms or molecules*.

* In the latter case it is actually the same atoms or molecules, which behave identically in all possible fields of force, that are identical with one another. Atoms or molecules differing in any way, for example, containing nuclei of different isotopes or being in different rotational states, must be considered as particles of quite different kinds.

Applying the point of view of the identity of particles in a consequential way, one has to renounce the numbering of particles. One can then no longer speak of "two states differing by the exchange of two particles" or of " $n!$ identical states differing by a permutation of n particles". We have to speak of "a state with an energy ϵ_k in which there are respectively two particles or n_k particles".

Instead of indicating the state of the entire gas by listing the particles in different energy states, one has to indicate the number of particles in each of these states, i.e. to indicate that there are

n_1 particles in the state with energy ϵ_1 ,

n_2 particles in the state with energy ϵ_2 .

Thus, the description of the state of a gas turns out to be less detailed than in classical statistics.

Since one cannot speak of a permutation of particles in a given state, the division by $n!$ makes no sense. Each state, irrespective of the number of particles which are in it, has the same statistical weight, namely the weight equal to unity.

The change in the method of calculating states leads to a basic change in the form of the statistical distribution. To obtain the latter we shall make use of the method of the preceding section. The mean number of particles in a state with an energy ϵ_k is given by the formula

$$\bar{n}_k = kT \frac{\partial}{\partial \mu} \ln \sum_{n_k} \left(\exp \frac{\mu - \epsilon_k}{kT} \right)^{n_k} \Omega(n_k). \quad (72.1)$$

Now, however, another value of $\Omega(n_k)$ must be substituted into (72.1). Since the state of a system containing an arbitrary number of particles $0 \leq n_k \leq \infty$ is non-degenerate, and the necessity of dividing by $n_k!$ disappears, for the number of states of a system containing n_k particles we have, instead of (71.3),

$$\Omega(n_k) = 1. \quad (72.2)$$

We have again assumed $\Delta\gamma = h^3$. Substituting this value of $\Omega(n_k)$ into (72.1), we find

$$\bar{n}_k = kT \frac{\partial}{\partial \mu} \ln \sum_{n_k} \left(\exp \frac{\mu - \epsilon_k}{kT} \right)^{n_k}. \quad (72.3)$$

The summation is carried out over the number of particles in the state with energy ϵ_k . In performing the summation it is necessary to distinguish between two kinds of particles, about which we have spoken in §1; particles which do not obey the exclusion principle, and those which do.

In the first case no restriction is imposed upon the number of particles which are in a given state. Their number can take on all integer values between zero and the total number of particles in the system. Thus,

$$\bar{n}_k = kT \frac{\partial}{\partial \mu} \ln \sum_{n_k=0}^N \left(\exp \frac{\mu - \epsilon_k}{kT} \right)^{n_k}. \quad (72.4)$$

Replacing the upper limit of the sum by infinity, we obtain

$$\bar{n}_k = kT \frac{\partial}{\partial \mu} \ln \sum_{n_k=0}^{\infty} \left(\exp \frac{\mu - \epsilon_k}{kT} \right)^{n_k}. \quad (72.5)$$

If the following inequality holds

$$\exp \frac{\mu - \epsilon_k}{kT} < 1, \quad (72.6)$$

then the sum in (72.5) represents an infinitely decreasing geometric progression and can easily be calculated. Namely,

$$\sum_{n_k=0}^{\infty} \left(\exp \frac{\mu - \epsilon_k}{kT} \right)^{n_k} = \left(1 - \exp \frac{\mu - \epsilon_k}{kT} \right)^{-1}$$

whence

$$\bar{n}_k = kT \frac{\partial}{\partial \mu} \ln \left(1 - \exp \frac{\mu - \epsilon_k}{kT} \right)^{-1} = \left(\exp \frac{\epsilon_k - \mu}{kT} - 1 \right)^{-1}. \quad (72.7)$$

Formula (72.7) gives the mean number of particles in the ideal gas which are in a state with energy ϵ_k , if the particles do not obey the exclusion principle. To this class of particles there belong atoms having zero spin, the molecules of saturated compounds also with zero spin and, in addition, as will be shown below, light quanta. The distribution (72.7) is called the Bose-Einstein distribution.

It should be noted that, since the sum (72.5) must always converge for any value of the energy, in particular for $\epsilon_k = 0$, in addition to the inequality (72.6)

$$e^{\mu/kT} < 1 \quad (72.8)$$

must also hold. The inequality (72.8) shows that the chemical potential for particles obeying the Bose–Einstein distribution must be an essentially negative quantity:

$$\mu < 0. \quad (72.9)$$

It should be recalled that in the Boltzmann distribution μ is also an essentially negative quantity, but is always very large in absolute value (see §60).

In the case of particles obeying the exclusion principle the number of particles n_k which can simultaneously be in an individual quantum state cannot exceed unity. Consequently, the possible values of n_k are restricted to two: $n_k = 0$ and $n_k = 1$. Substituting unity for the upper limit of the summation in formula (72.1), we have

$$\begin{aligned} \bar{n}_k &= kT \frac{\partial}{\partial \mu} \ln \sum_{n_k=0}^1 \left(\exp \frac{\mu - \epsilon_k}{kT} \right)^{n_k} = \\ &= kT \frac{\partial}{\partial \mu} \ln \left(1 + \exp \frac{\mu - \epsilon_k}{kT} \right) = \left(\exp \frac{\epsilon_k - \mu}{kT} + 1 \right)^{-1}. \end{aligned} \quad (72.10)$$

The distribution (72.10) represents the distribution of the particles of an ideal gas over states in the case where the particles obey the exclusion principle. The distribution (72.10) is called the Fermi–Dirac distribution.

In all cases encountered in practice the spacing between the energy levels of the translational motion is so small in comparison with the thermal energy kT that the energy spectrum can be assumed to be continuous. Then instead of the mean number of particles at the k th energy level it is necessary to introduce the mean number dn of particles with an energy lying between ϵ and $\epsilon + \delta\epsilon$. It is obvious that $dn = \bar{n} h^{-3} d\gamma$ where $h^{-3} d\gamma$ is the number of states corresponding to an energy in the interval $\epsilon, \epsilon + \delta\epsilon$. Substituting the mean value of the number of particles which are in one state from (72.7) into (72.10), we find

$$dn = \left(\exp \frac{\epsilon - \mu}{kT} \pm 1 \right)^{-1} \frac{d\gamma}{h^3}, \quad (72.11)$$

where the plus sign refers to the Fermi distribution, while the minus sign refers to the Bose distribution. The chemical potentials figuring in the Bose and Fermi distributions are determined from the normalization condition

$$\int dn = N, \quad (72.12)$$

which expresses the constancy of the number of particles in a given volume.

Comparing the derivation of the Bose distribution and the Fermi distribution with that of the Maxwell-Boltzmann distribution, we see first of all that the two distributions represent a realization of the Gibbs distribution for the case of an ideal gas whose particles obey the laws of quantum mechanics. There is a profound difference between the statistical distribution laws for particles obeying the laws of classical mechanics and for particles obeying the laws of quantum mechanics. This difference is not associated with any change in the statistical laws and not even with taking account of the discrete character of the energy spectrum, but with a radical change in the method of calculating the statistical weight of states. The difference between the methods for the calculation of statistical weights in classical statistics and the two quantum statistics is associated with the principle of identity of particles and is due to a profound difference between the behaviour of classical mechanical systems and the behaviour of atomic particles.

The difference between statistical weights in the Bose statistics and Fermi statistics is due solely to the difference between the laws of quantum mechanics obeyed by particles with an integer spin and those obeyed by particles with a half-integer spin. In this sense the often applied terminology "the Maxwell-Boltzmann classical statistics" or "the Fermi-Dirac quantum statistics and the Bose-Einstein quantum statistics" should be recognized as most inadequate. In reality the different forms of statistics are not in question but the different laws of quantum mechanics obeyed by the corresponding particles, i.e. the two forms of quantum mechanics: that for particles with an integer spin and that for particles with a half-integer spin. Statistical laws in all cases remain completely invariable. If the particles obey the laws of quantum mechanics for particles with an integer spin, the application of the laws of statistics leads to the Bose-Einstein distribution for the particles of an ideal gas. But if the particles obey the laws of quantum mechanics for particles with a half-integer spin, then the same statistics leads to the Fermi-Dirac distribution. Finally, if the particles obey the laws of classical mechanics, one obtains the Maxwell-Boltzmann distribution for a system of these particles.

The following question naturally arises. Experiment shows that the motion of atomic particles is described by the laws of quantum mechanics. Hence the behaviour of any ideal gas consisting of atomic particles must be described by one of the quantum distributions (72.7) or (72.10). Hence, is the Maxwell-Boltzmann distribution simply incorrect and not valid for real gases? A negative answer can be given to this question even without analysing the distributions (72.7) and (72.10), on the basis of general considerations. The laws of quantum mechanics are the laws of motion of particles including the laws of classical mechanics as a first approximation. Under certain conditions the laws of classical mechanics are a sufficiently good approximation; within the limits of this approximation it can be assumed that the motion of particles obeys the laws of classical mechanics. Consequently, there must exist also such conditions where the Maxwell-Boltzmann distribution reflects with a sufficient degree of accuracy the actual behaviour of ideal gases. We shall call gases obeying classical statistics non-degenerate gases. Conversely, ideal gases whose behaviour is governed by quantum laws are unified under the general name degenerate gases.

Our first problem is the discussion of the question of under which conditions a gas is degenerate and under which conditions it is non-degenerate or, in other words, when classical statistics can be considered to be applicable and when the laws of quantum statistics must be taken into account. In solving this problem we shall proceed from the fact that the Bose-Einstein distribution and the Fermi-Dirac distribution are more accurate laws. Hence use can be made of the Maxwell-Boltzmann law only when the difference between it and the quantum distributions (72.7) and (72.10) becomes sufficiently small.

Comparing the distributions (72.7), (72.10) and (71.6) we see that they have, in the general case, an essentially different character (fig. III.44). However, this difference vanishes if the following inequality is satisfied:

$$\exp \frac{\epsilon - \mu}{kT} \gg 1. \quad (72.13)$$

All three distributions have the same functional form when the inequality (72.13) is fulfilled. The lack of coincidence of the curves in fig. III.44 for large ϵ is associated with the fact that to these curves there correspond different μ . In this case the one in the denominator in (72.7) and (72.10) can be neglected, and the Bose and Fermi distributions automatically go over into the Maxwell-Boltzmann distribution. Thus, the inequality (72.13) is the condition of applicability of classical statistics.

For the fulfillment of the inequality (72.13) at energies $\epsilon \approx kT$ (for

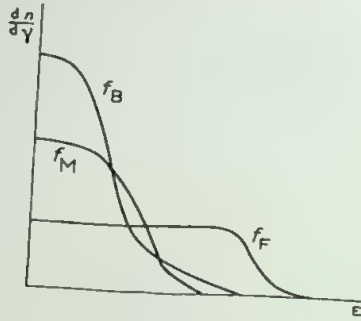


Fig. III.44

$\epsilon > kT$ the exponential rapidly increases) it is necessary that $e^{-\mu/kT} \gg 1$. Assume that this inequality is satisfied, so that for all energies

$$\left(\exp \frac{\epsilon - \mu}{kT} \pm 1 \right)^{-1} \approx \exp \frac{\mu - \epsilon}{kT}.$$

Then from the normalization condition (72.12) we find

$$\begin{aligned} N &= \int \exp \frac{\mu - \epsilon}{kT} \frac{dp_x dp_y dp_z dV}{h^3} = \\ &= \frac{2\pi(2m)^{3/2}}{h^3} V e^{\mu/kT} \int_0^\infty e^{-\epsilon/kT} \epsilon^{1/2} d\epsilon = \\ &= V \left(\frac{2\pi mkT}{h^2} \right)^{3/2} e^{\mu/kT}, \end{aligned}$$

whence

$$e^{-\mu/kT} = \frac{V}{N} \left(\frac{2\pi mkT}{h^2} \right)^{3/2}. \quad (72.14)$$

Thus, the criterion of the validity of the application of classical statistics is

$$e^{-\mu/kT} = \frac{V}{N} \left(\frac{2\pi mkT}{h^2} \right)^{3/2} \gg 1. \quad (72.15)$$

If the reverse inequality is fulfilled, then degeneracy arises and one cannot use the Maxwell—Boltzmann distribution.

We see that the criterion (72.15) contains several parameters. First of all, it contains the mass m of the particles; the larger the mass, the larger the left-hand side of the inequality. Further, (72.15) contains the density of the gas and its temperature T . As was to be expected, the inequality (72.15) is satisfied for high temperatures and is violated for low temperatures, so that at low temperatures quantum effects must appear. The fulfillment of the inequality (72.15) is also favoured by a small gas density.

In the inverse limiting case, when

$$\frac{V}{N} \left(\frac{2\pi mkT}{h^2} \right)^{\frac{3}{2}} \ll 1, \quad (72.16)$$

degeneracy of the gas arises. Thus, the degeneracy can be due to the following causes: (1) a small particle mass, (2) a large gas density, (3) a low temperature.

In order to estimate the order of magnitude of the quantities, we shall consider two numerical examples.

Suppose we have an electron gas. The mass of the electron is $m = 9.1 \times 10^{-28}$ g. We assume that the density of the electron gas is such that 1 cm^3 contains 6×10^{22} particles. Then it turns out that the condition of degeneracy is fulfilled up to temperatures of the order of 2000–3000 K.

In the case of atomic hydrogen, which is the lightest gas, degeneracy can arise only at very low temperatures and high densities, because the mass of a molecule of hydrogen is larger by a factor of 3700 than that of the electron. These temperatures and densities are considerably lower than those for which the interaction between atoms leading to condensation of the gas becomes important.

Thus, only in the case of an electron gas of high density can degeneracy occur at relatively high temperatures.

Another case of the quantum (degenerate) gas is the photon gas whose properties will be discussed in §76.

We shall not dwell in more detail here on the properties of the Bose and Fermi distributions, since it is more advisable to discuss them for real physical systems (the electron gas and the photon gas).

For all ordinary gases the difference between quantum statistics and classical statistics for not particularly large values of the temperature and density turns out to be negligibly small. The two quantum distributions, the Bose and Fermi distributions, can, with a high degree of accuracy, be re-

placed by the Maxwell distribution. The absence of any difference between Bose statistics and Fermi statistics can be understood if it is taken into account that the order of magnitude of the mean number \bar{n}_k of particles in an individual quantum state can be estimated by the following relations:

$$\bar{n}_k \approx e^{(\mu - \epsilon_k)/kT} \approx e^{\mu/kT} \ll 1.$$

In a non-degenerate gas for a high temperature and a small density of the gas the density of occupation of states is very small. In each state there is on the average much less than one particle, hence it is of no importance whether or not two or more particles can get into one state; even a pair will practically never in any event get into it.

In spite of the fact that classical statistics can always be applied to atomic gases (or, more precisely, quasi-classical statistics, since it is inevitable that we take into account discrete energy levels and introduce the factor $1/N!$), only the creation of quantum statistics made it possible to solve a number of the most important physical problems. Some of these will be discussed in following sections.

§73. Black-body radiation

The statistical theory of radiation played a very important role in the discovery of quantum theory. The classical electromagnetic theory of light, which explained a wide range of phenomena associated with the propagation of light and which was generally recognized by the end of the 19th century, encountered in the beginning of the 20th century insuperable difficulties in connection with the problem of the emission of light, in particular, with the problem of thermal radiation. We understand thermal radiation to be the whole of the radiation emitted by a heated body.

As is well known, the character of the light emitted, in particular its intensity as well as the dependence of the intensity on the frequency (the spectral composition of the radiation), is determined by the temperature and nature of the emitting body.

There are, however, cases where the spectral composition of the radiation does not depend on the nature of the emitter and is determined solely by its temperature. This is for so-called equilibrium radiation. Imagine a closed cavity with walls which do not conduct heat and which are maintained at a definite temperature T . The walls of the cavity will emit and absorb electromagnetic waves. Since all of the electromagnetic radiation is confined within

the closed cavity, a state of statistical equilibrium will be established in the system after a certain time. The walls of the cavity will emit as much electromagnetic energy per unit time as they absorb. A system of standing electromagnetic waves, stationary in time, will exist in the cavity.

The energy density of the corresponding electromagnetic field inside the cavity will be given by formula (12.6) of Part I:

$$\rho = \frac{E^2 + H^2}{8\pi}.$$

The thermal radiation will contain different frequencies. The energy density $\rho(\nu)$ in a given frequency interval $d\nu$ will, obviously, be different for different frequencies. The energy density of the radiation will also depend on the temperature T of the emitting walls. Thus,

$$\rho = \rho(\nu, T).$$

A simple thermodynamic consideration shows, however, that $\rho(\nu, T)$ does not depend on the nature of the emitter, in particular, on the nature of the walls (their absorption and emission properties, the state of the surface and so on).

Consider two cavities whose walls are heated to the same temperature but are made of different materials. We assume that the spectral energy density of the radiation depends on the nature of the emitter and is different in the two cavities. Then, by connecting the two cavities, one can disturb the state of equilibrium. Radiation will pass over to that cavity in which the density of radiation is smaller. As a result, the density will increase in this cavity, the walls of the cavity will absorb more radiation, and their temperature will increase. A difference between the temperatures of the walls of the two cavities will arise, which can be used for obtaining useful work.

The assumption which we have made leads to the conclusion of the possibility of spontaneous violation of the equilibrium state in a closed system and the possibility of constructing a perpetual motion machine of the second kind, which is, as is well known, impossible. Thus, it is proved that the spectral distribution of the energy density $\rho(\nu, T)$ of the equilibrium radiation is a universal function of the frequency ν and the temperature T .

A study of the absorption and emission properties of material bodies led Kirchhoff to the establishment of a very important theorem known as the Kirchhoff theorem.

The quantity $E(\nu)$, the energy emitted per cm^2 of the surface of a body

per second with a frequency between ν and $\nu + d\nu$ per unit frequency interval, will be called by us the radiative capacity of an arbitrary body.

Further, the part of the entire radiation energy with a frequency between ν and $\nu + d\nu$ incident on 1 cm^2 of the surface of a body which is absorbed inside the body * per unit frequency interval will be called by us the absorptive capacity of the body.

The Kirchhoff theorem reads that the ratio of the radiative capacity to the absorptive capacity $E(\nu)/A(\nu)$ is a universal function of the frequency and the temperature of the body but depends neither on the nature and properties of the body nor on its geometrical dimensions, i.e.

$$\frac{E(\nu)}{A(\nu)} = f(\nu, T). \quad (73.1)$$

It turns out that the universal function $f(\nu, T)$ is connected by a simple relation with the energy density of the equilibrium radiation $\rho(\nu, T)$ (T is the temperature of the body):

$$f(\nu, T) = \frac{c}{4\pi} \rho(\nu, T),$$

where c is the velocity of light. Thus, the Kirchhoff theorem can be written in the form

$$\frac{E(\nu)}{A(\nu)} = \frac{c}{4\pi} \rho(\nu, T).$$

The proof of the Kirchhoff theorem, which is of a very general character, can be found in any course on radiation theory **.

Since the absorptive capacity of a body can be found without any special difficulty from a measurement of the absorption coefficients and from geometrical considerations, finding the form of the function $\rho(\nu, T)$ appeared to be of great interest. From the Kirchhoff formula (73.1) it follows that a body with an absorptive capacity $A(\nu)$ equal to unity is of special importance.

* Not to be confused with the absorption coefficient, which characterizes the absorption of light per unit path in matter. The value of the absorptive capacity characterizes the absorption in the entire volume of the body.

** See, for example, M. Planck, *Vorlesungen über die Theorie der Wärmestrahlung* (Theory of thermal radiation) (Barth, Leipzig, 1913).

Such a body absorbs the entire electromagnetic energy incident on it for all frequencies. It is called an absolute black body.

For an absolute black body we have

$$E(\nu) = \frac{c}{4\pi} \rho(\nu, T). \quad (73.2)$$

Formula (73.2) shows that the absolute black body has a higher radiative capacity than all other bodies. Its radiative capacity is a universal function of the frequency ν and the temperature T .

By measuring the radiative capacity of an absolute black body, one can determine experimentally the form of the function $\rho(\nu, T)$.

Of course, not all bodies encountered in nature are absolute black ones. No matter what the nature of the surface of the body may be, a certain part of the radiative energy incident on it is reflected. However, the closed cavity filled with radiation which we have considered above is an absolute black body. Indeed, the entire radiation emitted by the walls of the cavity is also absorbed by them. If a small opening is made in the cavity, then by studying the spectral distribution of the radiation energy coming from this opening, one can find the function $\rho(\nu, T)$ experimentally. The size of the opening must be sufficiently small that radiation leakage through the opening does not lead to a fundamental departure from the equilibrium state. By means of such a model of an absolute black body the spectral energy distribution at different temperatures was investigated experimentally.

Fig. III.45 (see § 75) shows typical curves of this kind. The wavelength of the radiation is plotted on the horizontal axis, while the energy density of the radiation $\rho(\lambda, T)$ corresponding to a wavelength λ is plotted on the vertical axis. The energy density of radiation at a given wavelength is connected with $\rho(\nu, T)$ by the following relation:

$$\rho(\nu, T) d\nu = \rho(\lambda, T) d\lambda.$$

Taking into account that

$$d\nu = c \left| \frac{d\lambda}{\lambda^2} \right|,$$

we have

$$\rho(\lambda, T) = \frac{c}{\lambda^2} \rho(\nu, T).$$

Different curves in fig. III.45 refer to different temperatures. All the curves have a characteristic form. For long wavelengths the density of the radiation decreases with increasing λ . At a certain wavelength λ_{\max} it has a maximum, and then it again tends to zero in the direction of short wavelengths. The position of the maximum shifts in the direction of short wavelengths as the temperature increases.

§74. The classical theory of black-body radiation

We now pass on to the calculation of the spectral distribution function $\rho(\nu, T)$.

The electromagnetic radiation in a closed cavity forms a system of standing waves. We have considered such an electromagnetic field in §38 of Part I, where it has been shown that it can be replaced by a set of equivalent field oscillators. The energy of the field turned out to be equal to the sum of the energies of the oscillators. In the case of radiation in the cavity one has, corresponding to what has been said above, to ascribe to it a temperature which is equal to the temperature T of the radiating walls. We can therefore say that to each standing wave in the cavity there corresponds an oscillator with a frequency ν and an energy $\epsilon(\nu, T)$ which depends on the frequency as well as on the temperature T .

Every one of the oscillators which replace a system of standing waves can be in different states and can have a different energy $\epsilon(\nu, T)$. We shall, however, be interested not in the instantaneous but in the mean energy $\overline{\epsilon(\nu, T)}$ of the oscillators. Here the averaging is carried out over all possible states of the oscillator. The energy of the standing waves in unit volume of the cavity, whose frequencies lie between ν and $\nu + d\nu$, is numerically equal to the mean energy of all the oscillators which replace the normal oscillations and have frequencies in the same interval. If $g(\nu)d\nu$ is the number of oscillators, then the aforesaid can be written in the form

$$\rho(\nu, T)d\nu = \overline{\epsilon(\nu, T)} g(\nu)d\nu. \quad (74.1)$$

We have found the number of natural oscillations in §38 of Part I. In the case of electromagnetic waves it is necessary only to take into account that they are polarized and can have two directions of polarization. Formula (38.22) of Part I gives the number of oscillations with a frequency between ν and $\nu + d\nu$ for each direction of the polarization. For both directions of the polarization the number of oscillations must be doubled:

$$\rho(\nu, T) d\nu = \frac{8\pi\bar{\epsilon}}{c^3} \nu^2 d\nu. \quad (74.2)$$

The derivation of formula (74.2) did not involve any concepts of the quantum theory; it was obtained before the creation of the quantum theory. For the mean energy of the oscillator $\bar{\epsilon}$ its classical value

$$\bar{\epsilon} = kT$$

was taken, and the density of the equilibrium radiation at a temperature T was written in the form (the Rayleigh—Jeans law)

$$\rho(\nu, T) d\nu = \frac{8\pi kT}{c^3} \nu^2 d\nu. \quad (74.3)$$

The senselessness of formula (74.3) is obvious. It shows that the energy density of the electromagnetic field in a closed cavity increases monotonically with increasing frequency. Since oscillations of all frequencies, in particular $\nu \rightarrow \infty$, can arise in the cavity, formula (74.3) leads to an infinitely large energy density as $\nu \rightarrow \infty$:

$$E = \int_0^{\infty} \rho(\nu, T) d\nu \rightarrow \infty.$$

This result means that radiation sources confined in the cavity would have to emit radiation until the entire thermal energy contained in them went over into the radiation of the field and their temperature fell to absolute zero. Thus, for example, if a burning hot solid body placed in the cavity were the emitter, then from the result obtained it follows that equilibrium in the emitter—electromagnetic-field system would only be established after the hot body had cooled to absolute zero.

This conclusion has a simple meaning. According to the law of equipartition of energy, all degrees of freedom are equivalent and in an equilibrium state the same energy corresponds to each of them. The thermal energy contained in a crystal consisting of N atoms can be assumed to be distributed between $3N$ oscillators. The electromagnetic field in the cavity can also be considered as a set of oscillators. However, the number of these is immeasurably larger than $3N$. The wave numbers of possible standing waves in a closed cavity which has the form of a cube must satisfy the conditions

$$f_1 = \frac{\pi k_1}{L}, \quad f_2 = \frac{\pi k_2}{L}, \quad f_3 = \frac{\pi k_3}{L},$$

where L is the size of a face of the cube, and k_1, k_2, k_3 are numbers running over a series of integer values from zero to infinity. These conditions are equivalent to the conditions (38.11) for the crystal, but in the latter case the values of k_1, k_2 and k_3 are restricted by the number of particles N . Thus, the number of standing electromagnetic waves in the cavity and the corresponding number of electromagnetic field oscillators is infinitely larger than the number of oscillators needed for the description of thermal motion in a crystal. In an equilibrium state the entire energy must be contained in the field, since the same energy must correspond to each oscillator.

This result is in complete contradiction with experimental data. Experiment shows that the density of the thermal energy contained in the crystal is immeasurably larger than the density of the energy of the electromagnetic field. For example, at $T = 300$ K the density of thermal energy in a solid body turns out to be a factor of 10^{14} larger than the measured energy density inside a cavity with radiation. As to the spectral energy density distribution, expressed by formula (74.3), it turns out to be in agreement with the measured energy distribution in the black-body spectrum for small frequencies which satisfy the condition $h\nu \ll kT$. On the contrary, for large frequencies, when $h\nu \gg kT$, $\rho(\nu, T)$ increases with the frequency ν much more slowly than according to the law ν^2 .

Thus, the law of equipartition of energy leads to a complete disagreement of theory with experiment when applied to the problem of black-body radiation in the range of large frequencies. Historically, this was the first well investigated case of the inadequacy of classical concepts. The glaring contradiction with experiment to which classical statistics led, incited contemporary physicists to call the situation "the ultraviolet catastrophe". A way out of the contradiction was found by introducing the quantum theory.

§75. The Planck formula

The simplest, although not the most obvious method (from the physical standpoint) of finding the spectral distribution function $\rho(\nu, T)$ taking into account quantization consists of the following.

We substitute into formula (74.2) the value of the mean energy of a field oscillator calculated according to the theory of the quantum oscillator. We drop the zero energy of the oscillator $\frac{1}{2}h\nu$, choosing it to be the origin of energy. Then

$$\epsilon = nh\nu \quad (75.1)$$

and

$$\bar{\epsilon} = \frac{h\nu \sum_n n e^{-h\nu n/kT}}{\sum_n e^{-h\nu n/kT}} = \frac{h\nu}{e^{h\nu/kT} - 1}. \quad (75.2)$$

Substituting (75.2) into formula (74.2), we find the following expression for the mean energy of the vacuum electromagnetic field per unit volume for a frequency lying between ν and $\nu + d\nu$:

$$\rho(\nu, T)d\nu = \frac{8\pi h\nu^3 d\nu}{c^3(e^{h\nu/kT} - 1)}. \quad (75.3)$$

Formula (75.3) is called the Planck formula. This formula was first derived semi-empirically, since formula (75.2) for the energy of an oscillator was unknown. Formula (75.3) and the Planck constant h contained in it were found from experiment.

In the two limiting cases, $h\nu/kT \ll 1$ and $h\nu/kT \gg 1$, the Planck formula is simplified. In the first case

$$e^{h\nu/kT} \approx 1 + \frac{h\nu}{kT}$$

and formula (75.3) reduces to the form

$$\rho(\nu, T)d\nu \approx \frac{8\pi kT}{c^3} \nu^2 d\nu, \quad (75.4)$$

i.e. goes over into the classical formula (74.3) for the mean energy density of black-body radiation.

For $h\nu/kT \gg 1$

$$(e^{h\nu/kT} - 1)^{-1} \approx e^{-h\nu/kT},$$

so that

$$\rho(\nu, T)d\nu \approx \frac{8h\nu^3}{c^3} e^{-h\nu/kT} d\nu. \quad (75.5)$$

This last formula is called Wien's law.

Going from $\rho(\nu, T)$ to the spectral distribution of the radiation density with respect to wavelength, $\rho(\lambda, T)$, we can write formulae (75.3), (75.4) and (75.5) in the following forms:

$$\rho(\lambda, T) = \frac{8\pi hc}{\lambda^5} (e^{hc/kT\lambda} - 1)^{-1}, \quad (75.6)$$

$$\rho(\lambda, T) = \frac{8\pi kT}{\lambda^4} \left(\frac{hc}{kT\lambda} \ll 1 \right), \quad (75.7)$$

$$\rho(\lambda, T) \approx \frac{8\pi hc}{\lambda^5} e^{-hc/kT\lambda} \left(\frac{hc}{kT\lambda} \gg 1 \right). \quad (75.8)$$

The curves corresponding to formula (75.6) are shown in fig. III.45. For large wavelengths $\rho(\lambda, T)$ decreases with increasing wavelengths as λ^{-4} , while for small wavelengths $\rho(\lambda, T)$ tends to zero as $\lambda^{-5} e^{-hc/kT\lambda}$. The function $\rho(\lambda, T)$ has a maximum at a wavelength λ_{\max} which can be found from the condition

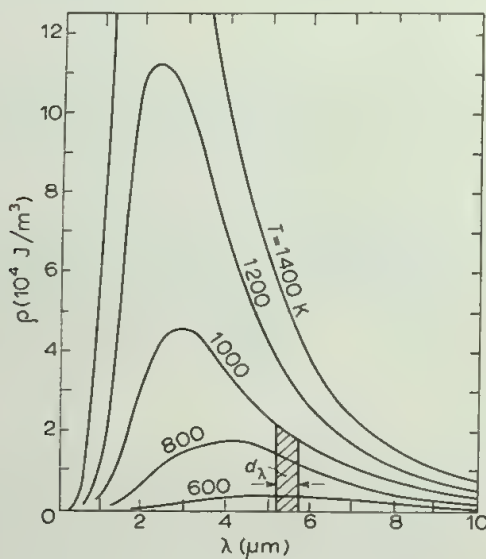


Fig. III.45

$$\frac{\partial \rho(\lambda, T)}{\partial \lambda} = 0$$

or

$$-\frac{5}{\lambda^6} \frac{1}{e^{hc/kT\lambda} - 1} + \frac{hc}{\lambda^7 kT} \frac{e^{hc/kT\lambda}}{(e^{hc/kT\lambda} - 1)^2} = 0.$$

Denoting the quantity $hc/kT\lambda_{\max}$ by x , we can write the last equation in the form

$$\frac{x e^x}{e^x - 1} = 5.$$

The solution of this transcendental equation gives

$$x = 4.96, \quad \lambda T \approx 2 \times 10^{-3} hc/k. \quad (75.9)$$

Formula (75.9) shows that the position of the maximum of the energy density of black-body radiation shifts in the direction of small wavelengths with increasing temperature. This is the so-called displacement law.

The value of the quantum constant h can be determined from the displacement law. After the proper choice of this value the Planck formula turns out to be in excellent agreement with experimental data.

§ 76. The statistics of the photon gas

As we have already pointed out in the introductory chapter, contemporary quantum theory, in accord with experimental data, states that radiation possesses corpuscular as well as wave properties. Although from the point of view of ordinary ideas it is impossible to combine the properties of a wave and a particle in one object, one sometimes has to make use of the wave aspect and sometimes the corpuscular aspect in order to explain different optical phenomena. Thus, for example, in the phenomena of interference or diffraction the wave nature of radiation is manifested, whereas in the photoelectric effect or the scattering of hard X-rays the corpuscular nature of radiation is revealed. From the corpuscular point of view, radiation can be considered as a flux of light quanta, or photons, moving in space with the velocity of light c . Photons arise and vanish when light is emitted and

absorbed by atoms respectively. Their energy is equal to $\epsilon = \Delta E$, where ΔE is the difference between the energy levels of the emitting system.

All photons move in vacuum with the same velocity, but different photons can have different energy and different momentum. The energy and momentum of photons are related by the expression

$$\epsilon = pc, \quad (76.1)$$

which is a general formula relating these quantities for any object moving with the velocity of light. The energy and momentum of a photon depend on the frequency according to the formulae

$$\epsilon = h\nu = \hbar\omega, \quad (76.2)$$

$$p = h\nu/c = \hbar\omega/c. \quad (76.3)$$

Like other material particles, photons also possess angular momentum (see § 1, as well as Part V).

It turns out that as an emitting system (an atom or a molecule) emits radiation its angular momentum must decrease by an amount which is a multiple of \hbar . The corresponding angular momentum is carried away by the outgoing photon. Thus, the angular momentum expressed in units of \hbar is an integer. Like all other particles with an integer angular momentum, photons obey Bose-Einstein statistics.

From the corpuscular point of view equilibrium radiation filling a closed cavity must be considered as a photon gas filling the volume V of the container. The particles of the photon gas are moving randomly in all directions in the container, and their directions of flight change in collisions with the walls of the container. There is no interaction between the photons. Hence the photon gas must in its properties be similar to an ordinary molecular ideal gas filling a closed container.

However, in addition to the similarity between the photon gas and the molecular gas there is also a very profound difference between them. The most fundamental difference between the photon gas and the molecular gas is the fact that for the photon gas one cannot speak of a fixed number of particles. In contrast to ordinary particles (electrons, protons or atoms), photons can arise or vanish at the moment of emission or absorption of light by atoms. Hence the number of photons in the cavity cannot be considered as fixed.

Another difference between photons and gas molecules is the fact that the

former all move with the same speed. In fact, however, this property of the photon gas is not associated with the specific nature of photons. For very large values of the kinetic energy of all particles their velocities approach the velocity of light and the differences between the velocities of individual particles are progressively smoothed over. Hence this difference between the photon gas and a molecular gas is not essential. Of importance only is the fact that in the photon gas, as well as in a molecular gas, there is a certain distribution of particles over momenta and energies.

Finally, there is one more difference between the photon gas and a gas of ordinary particles, a difference of principle rather than of a practical character. As has been shown in §6 and §16, the establishment of the velocity distribution (or the momentum distribution) of molecules is closely associated with the interaction occurring between them in molecular collisions. But photons do not collide with each other at all. An equilibrium distribution of photons can be established only in the case where there is a body in the cavity which can absorb and emit photons. In the process of absorption and subsequent emission photons of one frequency transform into photons of other frequencies. In this case the number of photons does not remain constant, although their total energy must be conserved. In particular, the walls of a cavity containing a "photon gas" can act as such a body.

We shall now show that, proceeding from the concept of a photon gas, one can arrive at the Planck formula with the same success as using the wave concept.

From the corpuscular point of view the function $\rho(\omega, T)$ can be interpreted in the following way. In the frequency interval $\omega, \omega + d\omega$, or in the corresponding energy interval of the photons $\epsilon, \epsilon + d\epsilon$, let there be $d\Omega$ quantum states of photons per unit volume. Further, let the mean number of photons in each state be equal to $\bar{n}(\epsilon)$. Then the mean number of photons with an energy between ϵ and $\epsilon + d\epsilon$ is equal to

$$d\bar{n} = \bar{n}(\epsilon)d\Omega. \quad (76.4)$$

Their mean energy is equal to $\epsilon\bar{n}(\epsilon)d\Omega$. But this energy represents none other than the energy density of radiation with a frequency between ω and $\omega + d\omega$. Thus,

$$\rho(\epsilon, T)d\omega = \epsilon\bar{n}(\epsilon)d\Omega. \quad (76.5)$$

Our problem is to find $\bar{n}(\epsilon)$, i.e. the mean number of particles in a gas of a variable number of particles. Taking into account the fact that photons are

particles with spin equal to unity, then for $\bar{n}(\epsilon)$ one can write the Bose-Einstein distribution.

It is necessary, however, to take into account the peculiar feature of the photon gas which is associated with the possibility of absorption and emission of photons by the walls of the container or by material bodies placed inside the cavity.

The number of particles in a photon gas is variable and depends on the state of the gas. Hence, in contrast to an ordinary molecular gas, the free energy of a photon gas depends not only on the variables V and T but also on the number of particles N in the gas. For a given value of V and T the number of photons in an equilibrium state will have a value N_0 such that the free energy $F(V, T, N_0)$ has its minimum value. Thus, it can be said that an equilibrium state of a photon gas occurs when the following equality is fulfilled:

$$\frac{\partial F(V, T, N)}{\partial N} = \mu(V, T) = 0. \quad (76.6)$$

Here we have made use of formula (60.3').

Eq. (76.6) shows that the chemical potential of an equilibrium photon gas is zero.

Thus, the mean number of particles of a photon gas per unit volume which have an energy ϵ , must, by virtue of (72.7) and (76.6), be written in the form

$$\bar{n}(\epsilon) = \frac{1}{e^{\epsilon/kT} - 1}. \quad (76.7)$$

The number of photons with an energy between ϵ and $\epsilon + d\epsilon$ is equal to

$$d\bar{n}(\epsilon) = \frac{2}{e^{\epsilon/kT} - 1} \frac{4\pi p^2 dp}{h^3} = \frac{8\pi}{h^3 c^3} \frac{\epsilon^2 d\epsilon}{e^{\epsilon/kT} - 1}. \quad (76.8)$$

The factor 2 is introduced to take into account the two-fold degeneracy of the states of photons with a given momentum p . For a given value of p there are two states corresponding to the two possible independent directions of polarization of light.

The total number of photons in the equilibrium radiation can be found by integrating (76.8) over all values of ϵ .

Making use of formula (76.2), ϵ can be expressed in terms of the frequency ω . Performing this substitution and passing over from the number of photons to their energy, we find

$$\begin{aligned}\rho(\omega, T)d\omega &= \epsilon \bar{n}(\epsilon)d\Omega = \\ &= \frac{8\pi}{h^3 c^3} \frac{\epsilon^3 d\epsilon}{e^{\epsilon/kT} - 1} = \frac{\hbar \omega^3 d\omega}{\pi^2 c^3 (e^{\hbar \omega/kT} - 1)},\end{aligned}\quad (76.9)$$

i.e. the Planck formula.

It should be stressed that if the Boltzmann distribution and not the Bose–Einstein distribution were applied to photons, then instead of the Planck formula one would obtain formula (75.5) which is valid only when $\hbar\omega/kT \gg 1$. Indeed, substituting the expression $\bar{n} = e^{-\epsilon/kT} = e^{-\hbar\omega/kT}$ for $\bar{n}(\epsilon)$ from (76.7), we obtain Wien's law (75.5) instead of the Planck formula (76.9).

Thus, classical statistics cannot be applied to photons. The region of applicability of classical statistics to the photon gas is restricted by the condition $\hbar\omega/kT \gg 1$. This condition is the reverse of that for classical statistics to be applicable to electromagnetic field oscillators. Thus, for high frequencies (or low temperatures) it is the corpuscular properties of radiation that are predominant, whereas for low frequencies (or high temperatures) it is the wave properties of radiation that are predominant.

The energy of electromagnetic radiation or the energy per unit volume of a photon gas is obtained from (76.9) by integrating over all frequencies:

$$u = \int_0^\infty \rho(\omega, T)d\omega = \frac{\hbar}{\pi^2 c^3} \int_0^\infty \frac{\omega^3 d\omega}{e^{\hbar\omega/kT} - 1}.$$

To calculate the integral we introduce a new variable, $x = \hbar\omega/kT$. Then

$$u = \frac{\hbar}{\pi^2 c^3} \left(\frac{kT}{\hbar}\right)^4 \int_0^\infty \frac{x^3 dx}{e^x - 1}.$$

The value of this integral is obtained in Appendix IV:

$$\int_0^\infty \frac{x^3 dx}{e^x - 1} = \frac{\pi^4}{15}.$$

Hence finally we have

$$u = \frac{\pi^3 k^4 T^4}{15(hc)^3} = \alpha T^4. \quad (76.10)$$

The energy density of black-body radiation turns out to be proportional to the fourth power of the absolute temperature (the Stefan—Boltzmann law).

The constant α figuring in (76.10) contains only the universal constants h , c and k . The Stefan—Boltzmann law is widely applied in heat engineering for the calculation of the radiative capacity of heated surfaces. Although radiators encountered in practice are not black bodies, the application of the Stefan—Boltzmann law leads to good results for all solid radiators except metals. For the latter the emitted energy increases as a higher power of the temperature.

The total energy of radiation in a volume V is equal to

$$E = \alpha VT^4. \quad (76.11)$$

Further, we find the free energy of black-body radiation. According to formula (30.12), we have

$$F = -T \int \frac{E dT}{T^2} = -\frac{1}{3} \alpha VT^4. \quad (76.12)$$

The entropy of radiation is

$$S = -\partial F / \partial T = \frac{4}{3} \alpha VT^3. \quad (76.13)$$

And the radiation pressure p is equal to

$$p = -\partial F / \partial V = \frac{1}{3} \alpha T^4 = \frac{1}{3} E / V. \quad (76.14)$$

Radiation pressure was first discovered by P.N. Lebedev. This discovery was of fundamental importance. It allowed proof of the impossibility of constructing a perpetual motion machine of the second kind in which radiation could be used as the working medium.

Radiation pressure, which is very small in terrestrial conditions, assumes an extremely great importance in astrophysics. As is shown by formula (76.14), the radiation pressure increases very rapidly with increasing temperature. At very high temperatures, such as occur in astrophysical conditions, the radiation pressure turns out to be larger than the gas pressure, and plays a basic role in a number of astrophysical processes.

Finally, a simple calculation shows that the thermodynamic potential Φ of radiation is equal to zero:

$$\Phi = F + pV = 0.$$

This is in agreement with our requirement $\mu = 0$ for the photon gas.

§77. The properties of liquid helium II

A very interesting example of a macroscopic system in which quantum effects are displayed is liquid helium II, the only system which remains liquid down to absolute zero. All other liquids solidify at temperatures which are too high for quantum effects to be shown.

As shown by experiment, liquid helium can exist in two modifications which are called liquid helium I and liquid helium II and which differ sharply from each other in physical properties.

Fig. III.37 (see §62) shows the phase diagram of helium. It is seen that at pressures above 30 atm liquid helium I, which represents the high-temperature modification, goes over into the solid state as the temperature decreases. However, at pressures below 30 atm helium does not solidify at any temperature and remains liquid down to $T = 0$ K. On curve II the phase transition of helium I into the other modification takes place. The behaviour of the heat capacity (fig. III.46), density and a number of other properties of helium depending on the temperature is indicative of this transition. The heat capacity undergoes a jump at the transition point; on the density curve a break is observed at this point, and so on. Since the latent heat of the phase

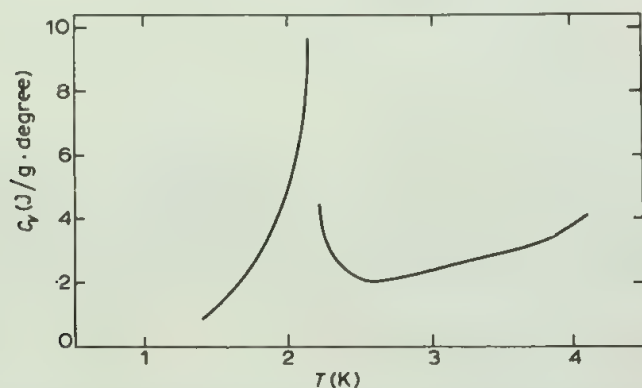


Fig. III.46

transition helium I — helium II is zero, this phase transition is a typical phase transition of the second kind.

Liquid helium II possesses a number of remarkable properties which are due to its quantum nature. Some of these will be described below*.

L.D.Landau proposed a statistical theory of liquid helium II based on certain assumptions about the character of the energy spectrum of this system. New studies on the theory of helium II will be discussed in Part V. In these studies the assumptions which underlie the theory of Landau are derived from the general propositions of the quantum mechanics of a system of particles.

Consider a certain amount of liquid helium II confined in a container. The liquid as a whole represents a quantum system; its possible energy values form a certain energy spectrum. At very low temperatures the discrete character of the energy spectrum of the liquid cannot be disregarded, in spite of the fact that the liquid is a macroscopic system. We have to determine the character of the energy spectrum of the macroscopic quantum system for very small excitation energies, when the system can only be at energy levels close to the ground level which the system is in at absolute zero.

An accurate calculation of the energy levels of a system consisting of a large number of strongly interacting particles appears to be impossible at present. This holds equally for liquid helium II, crystals, mutually interacting electrons and all other systems of interacting particles. Nevertheless, it is possible to establish certain general properties of the energy spectrum of such systems for small excitation energies. In particular, such an energy spectrum will be possessed by liquid helium II, in which the smallness of the excitation energy is ensured by the temperature.

The basic property of the energy spectrum of any macroscopic system for small excitation energies is the fact that the excitation energy can be resolved into a set of independent "elementary excitations".

We shall consider for concreteness the energy spectrum of the elastic oscillations of a crystal or a liquid for small excitation energies. We have seen before that the motion of the atoms of a solid body can be resolved into independent elastic waves which do not interact with each other and are propagating over the entire volume of the body. The only difference between a crystal and a quantum liquid is the fact that in the former both longitudinal and transverse waves can propagate, whereas in the liquid only longitudinal waves (compression and expansion waves) can exist. Each of these waves

* See the book which we follow: W.H.Keesom, *Helium* (Elsevier Publ. Co., Amsterdam, 1942, repr. 1959).

carries a definite invariable energy which can be considered as an elementary excitation. The energy of the entire body can be considered as a set of elementary excitations, i.e. as the sum of the energies of all independent elastic waves propagating in the body. From the above it is clear that an elementary excitation corresponds to an excitation energy of the body as a whole and can in no way be referred to an individual atom with an excess energy over other atoms in the body. Each of the elementary excitations representing a sound wave moves in the body, undergoes reflections from its walls, moves in a new direction and so on. An elementary excitation possesses a definite momentum in addition to an energy.

The motion of all elementary excitations in a body can be likened to the motion of non-interacting quasi-particles, that is excitation quanta, which form an ideal gas inside the body. A complete analogy can be drawn between light waves and light quanta on the one hand, and between elastic waves and excitation quanta in a crystal on the other hand. Just as the light field can be treated as a set of light quanta (photons), the field of elastic waves which fill a crystal can be replaced by a gas of excitation quanta which are often called phonons.

However, an essential reservation should be made. This analogy, which is very convenient for carrying out a number of calculations, has only a formal character. Sound quanta have no direct physical reality and serve only for a mathematical expression of the properties of a discrete set of elastic waves in a crystal. Keeping in mind this reservation, we shall assume in what follows that the excitation energy of a body represents the energy of excitation quanta which fill its entire volume like an ideal gas. The energy of an excitation quantum ϵ has, in the general case, an unknown functional relationship with its momentum p .

We shall now consider the energy spectrum for the case of liquid helium * in more detail. The basic properties of liquid helium can be derived from certain simple assumptions about the form of the spectrum. That is, it should be assumed that in helium II there are two forms of excitation quanta: long-wave and short-wave ones. The first, which have a large wavelength λ , carry a small momentum $p = h/\lambda$ and a small energy $\epsilon(p)$. For small p the function $\epsilon(p)$ can be expanded in a series in powers of p , and one can write

$$\epsilon \approx \text{const} \cdot p. \quad (77.1)$$

* We shall return to the problem of the energy spectrum of helium II in Part V. The problem of collective excitations will be discussed in more detail in Part VI.

Long-wave excitations in liquid helium II represent elastic longitudinal compression and expansion waves. Therefore the constant in (77.1) is simply the velocity c of propagation of sound waves. Thus, for long-wave quanta one can write

$$\epsilon = cp. \quad (77.2)$$

In hydrodynamics it turns out that, when sound waves of a small amplitude arise in a liquid, then the liquid comes into a state of vortex-free (potential) motion. However, in the general case the flow of a non-ideal (viscous) liquid is a vortex motion. Hence in addition to longitudinal sound waves in liquid helium II there must exist other elementary excitations. We shall assume that besides long-wave sound excitation quanta there exist also in helium II short-wave excitation quanta whose wavelength is close to a certain wavelength λ_0 . The corresponding momentum of the short-wave quanta is close to $p_0 = h/\lambda_0$. We assume that the energy of quanta with momentum p_0 has a minimum value in comparison with the energy of all quanta whose momenta are close to p_0 . In other words, we assume that the energy of elementary excitations has the form of the curve shown in fig. III.47. It can then be said that in the liquid besides long-wave excitation quanta whose

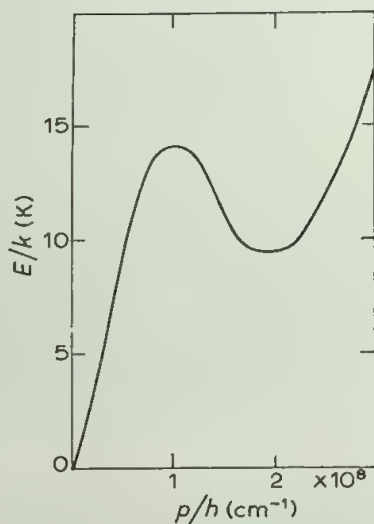


Fig. III.47

momentum is close to zero there will also exist quanta with a momentum $p \approx p_0$. The energy of such quanta can be written in the form

$$\epsilon \approx \epsilon(p_0) + \frac{(p-p_0)^2}{2\mu}, \quad (77.3)$$

where $\epsilon(p_0)$ and μ are constants whose values must be determined experimentally. In expansion (77.3) in powers of $p - p_0$ the term proportional to the first power of this difference is absent, because, by the assumption, $\epsilon(p)$ has a minimum at the point $p = p_0$. The constant in the second term is denoted by μ , to stress that the energy of short-wave quanta formally looks the same as the energy of ordinary particles.

Of course, there are no grounds for assuming beforehand that in the excitation spectrum of the quantum liquid there exist mainly just quanta of the two types quoted. However, the introduction of such a spectrum is justified by the fact that by means of it, it turns out to be possible to give a quantitative explanation of all the peculiar phenomena which occur in helium II.

At the same time it should be kept in mind that in addition to short-wave and long-wave excitation quanta in the liquid there are also quanta of intermediate wavelengths, but that the number of such quanta is relatively small. As we have just stressed, excitation quanta move over the entire volume of the body without interacting with each other (for small excitations), like the particles of an ideal gas filling the volume of the body. Long-wave excitation quanta can be likened to photons, while short-wave quanta behave as ordinary particles of an ideal gas which have a mass μ .

To avoid any misunderstanding, we stress once more that this analogy has only a mathematical nature. In reality, each excitation represents a particular form of motion of all the atoms of the liquid. Hence a short-wave excitation cannot be represented as a real particle moving in the liquid. However, the mathematical analogy between a set of excitation quanta and an ideal gas allows one to find the thermodynamic functions of liquid helium easily.

§ 78. Statistical theory of liquid helium II

The presence of thermal excitation quanta in liquid helium II means, from the macroscopic point of view, that it has a free energy F which can be assumed to be made up of the free energy due to the existence of long-wave excitations and the free energy due to the existence of short-wave excitations:

$$F = F_{Lw} + F_{s.w} . \quad (78.1)$$

We shall write an expression for each of the terms separately.

The free energy of long-wave quanta F_{Lw} can immediately be written in analogy with the free energy of a solid body at low temperature, taking into account that now only longitudinal waves can exist, whereas transverse waves are absent. Thus,

$$3N = \frac{4\pi V \nu_{\max}^3}{3c^3} , \quad (78.2)$$

where N is the number of atoms of the liquid in a volume V , ν_{\max} is the maximum frequency of the sound waves, and c is the velocity of sound.

Hence, substituting the value $\theta_c = h\nu_{\max}/k$ into (53.7) and taking into account (78.2), we find for F_{Lw}

$$F_{Lw} = - \frac{4}{45} \frac{\pi^5 (kT)^4 V}{h^3 c^3} . \quad (78.3)$$

The calculation of $F_{s.w}$ is somewhat more complicated. Short-wave quanta behave like the particles of an ideal gas. Their energy, which is determined by formula (77.3) can be assumed to be large in comparison with kT at sufficiently low temperatures where it is still possible to speak of independent elementary excitations. For this the following inequality must in any case be fulfilled:

$$\epsilon(p_0) \gg kT .$$

We shall see below that this inequality is indeed fulfilled in liquid helium II. Hence the short-wave quantum distribution function has the form of the classical Boltzmann distribution. The free energy of a classical ideal gas (taking into account the identity of particles) has the form

$$F_{s.w} = -N_{s.w} kT \ln \left(\frac{eV}{N_{s.w}} \right) \int e^{-\epsilon/kT} \frac{d\gamma}{h^3} , \quad (78.4)$$

where $d\gamma = dp_x dp_y dp_z$ and $N_{s.w}$ is the number of short-wave excitation quanta. The value of $N_{s.w}$ is, however, not a fixed quantity, but depends on the temperature of the liquid. It increases with increasing excitation, i.e.

with increasing temperature of the liquid *. For a given temperature the number of short-wave excitation quanta is determined from the condition of the minimum free energy:

$$\frac{\partial F_{s.w}}{\partial N_{s.w}} = 0. \quad (78.5)$$

Substituting (78.4) into the condition (78.5), we find for the number of short-wave quanta the expression

$$N_{s.w} = V \int e^{-\epsilon/kT} \frac{d\gamma}{h^3}. \quad (78.6)$$

Substituting $N_{s.w}$ into (78.4), we find for the free energy

$$F_{s.w} = -kTV \int e^{-\epsilon/kT} \frac{d\gamma}{h^3}. \quad (78.7)$$

We calculate the integral contained in (78.7). Obviously, we have

$$\begin{aligned} \int e^{-\epsilon/kT} \frac{d\gamma}{h^3} &= 4\pi \int \exp\left(-\frac{\epsilon(p_0)}{kT}\right) \exp\left(-\frac{(p-p_0)^2}{2\mu kT}\right) \frac{p^2 dp}{h^3} = \\ &= 4\pi \exp\left(-\frac{\epsilon(p_0)}{kT}\right) \int \exp\left(-\frac{(p-p_0)^2}{2\mu kT}\right) \frac{p^2 dp}{h^3}. \end{aligned}$$

The range of integration with respect to the momentum of the excitation quanta is not accurately determined. Since, however, the integrand rapidly decreases with increasing difference $p - p_0$, and for $(p-p_0)^2/2\mu \gg kT$ practically reduces to zero, the range of integration can be extended to $\pm\infty$. Then we have

$$\int e^{-\epsilon/kT} \frac{d\gamma}{h^3} = 4\pi \exp\left(-\frac{\epsilon(p_0)}{kT}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(p-p_0)^2}{2\mu kT}\right) \frac{p^2 dp}{h^3}.$$

* The dependence of $N_{s.w}$ on the temperature allows us once again to convince ourselves that the treatment of elementary excitations as quasi-particles has a relative character.

Since the integrand reduces to zero when $(p-p_0)^2/2\mu \gg kT$, the slowly varying function p^2 can be brought out of the integral sign and taken at the point $p = p_0$. Then

$$\begin{aligned} \int e^{-\epsilon/kT} \frac{d\gamma}{h^3} &= 4\pi p_0^2 \exp\left(-\frac{\epsilon(p_0)}{kT}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(p-p_0)^2}{2\mu kT}\right) \frac{dp}{h^3} = \\ &= \frac{4\pi p_0^2 (2\mu kT)^{\frac{1}{2}}}{h^3} \exp\left(-\frac{\epsilon(p_0)}{kT}\right). \end{aligned} \quad (78.8)$$

Thus, we obtain finally

$$F_{s.w} = - \frac{4\pi (2\pi\mu)^{\frac{1}{2}} (kT)^{\frac{3}{2}} V p_0^2}{h^3} \exp\left(-\frac{\epsilon(p_0)}{kT}\right), \quad (78.9)$$

$$N_{s.w} = \frac{4\pi p_0^2 V (2\pi\mu kT)^{\frac{1}{2}}}{h^3} \exp\left(-\frac{\epsilon(p_0)}{kT}\right). \quad (78.10)$$

Substituting the values of $F_{l.w}$ and $F_{s.w}$ from (78.3) and (78.9) into (78.1), we find the expression for the free energy of liquid helium II:

$$F = - \frac{4}{45} \frac{\pi^5 (kT)^4 V}{h^3 c^3} - \frac{4\pi (2\pi\mu)^{\frac{1}{2}} (kT)^{\frac{3}{2}} V p_0^2}{h^3} \exp\left(-\frac{\epsilon(p_0)}{kT}\right). \quad (78.11)$$

Correspondingly, the entropy and heat capacity of helium II are equal to

$$S = \frac{16}{45} \frac{\pi^5 k^4 T^3 V}{h^3 c^3} + \frac{4\pi (2\pi\mu)^{\frac{1}{2}} p_0^2 k^{\frac{3}{2}} T^{\frac{1}{2}} V}{h^3} \left\{ \frac{3}{2} + \frac{\epsilon(p_0)}{kT} \right\} \exp\left(-\frac{\epsilon(p_0)}{kT}\right), \quad (78.12)$$

$$\begin{aligned} C_V &= \frac{16}{15} \frac{\pi^5 k^4 T^3 V}{h^3 c^3} \\ &+ \frac{4\pi (2\pi\mu)^{\frac{1}{2}} p_0^2 k^{\frac{3}{2}} T^{\frac{1}{2}} V}{h^3} \left\{ \frac{3}{4} - \frac{\epsilon}{2kT} + \frac{\epsilon^2}{k^2 T^2} \right\} \exp\left(-\frac{\epsilon(p_0)}{kT}\right). \end{aligned} \quad (78.13)$$

We see that all the thermodynamic quantities are made up of two parts which are due respectively to the long-wave and the short-wave excitations. The first part varies with temperature according to the same power law as in the case of crystals. The second part depends on the temperature exponentially, i.e. proportional to $\exp[-\epsilon(p_0)/kT]$. The values of the constants were

determined from the measurements of the entropy and heat capacity of helium II, and turned out to be equal to

$$\epsilon(p_0)/k = 9.6 \text{ K}, \quad p_0/h = 12.25 \times 10^8 \text{ cm}^{-1}, \quad \mu = 0.75 m_{\text{He}}.$$

For these values of the constants the power part of the heat capacity and entropy exceeds the exponential part at temperatures which are lower than about 1 K. On the contrary, at high temperatures the exponential (short-wave) part predominates. The temperature trend of the thermodynamic quantities is in complete agreement with experiment.

The most outstanding property of liquid helium II, which was discovered by P.L.Kapitza, is its "superfluidity". Namely, measurements of the viscosity of helium II flowing through fine slits and capillary tubes showed the viscosity to be negligibly small. Owing to this helium II flows practically without obstruction through the finest capillary tubes. A consequence of the superfluidity of helium II is its extraordinarily high thermal conductivity ("the thermal superconductivity"), which was discovered experimentally before the superfluidity. Owing to the negligibly small viscosity, characteristic convection streams arise in helium II, which make it possible to transfer considerable amounts of heat under conditions in which an ordinary viscous liquid devoid of convection agitation has a negligibly small thermal conductivity. The phenomenon of superfluidity finds a complete explanation in the theory given above. It turns out to be closely associated with the character of the energy spectrum of helium II.

Let us consider the flow of helium II along a solid wall. For convenience of the treatment we shall transform to the reference frame in which the helium is at rest and the solid wall is moving. From the point of view of excitation quanta any process of energy dissipation due to the viscosity can be considered in the following way. In the reference frame which we have chosen the energy of the helium is initially given and is determined by the number of elementary thermal excitations. The interaction between the wall and the liquid leads to the appearance of an additional internal motion in the liquid layer adjacent to the wall. This internal motion represents the thermal motion of the particles of the liquid. Thus, the energy dissipation consists in the appearance of excitation quanta (thermal motion) in the liquid. We shall assume in the beginning that in the helium II there were initially no excitation quanta, i.e. that its temperature T was equal to zero. Let an excitation quantum with a momentum p and an energy $\epsilon(p)$ arise in the helium. Then the internal energy of the helium will become equal to $\epsilon(p)$. In the reference frame in which the helium flows while the wall is at rest, its

energy, according to the rules for the transformation of energy in a relative motion, is equal to

$$E = \frac{1}{2}mv^2 + \epsilon(p) + \mathbf{p} \cdot \mathbf{v}, \quad (78.14)$$

where v is the velocity of flow, $\frac{1}{2}mv^2$ is the kinetic energy of the liquid, and $\epsilon(p) + \mathbf{p} \cdot \mathbf{v}$ is the change in its energy.

When energy is dissipated the kinetic energy of the flowing liquid can only decrease, i.e. $\epsilon + \mathbf{p} \cdot \mathbf{v} < 0$. The smallest value of the quantity $\epsilon + \mathbf{p} \cdot \mathbf{v}$ is reached when a quantum with momentum \mathbf{p} directed antiparallel to \mathbf{v} arises. It is equal to $\epsilon - pv$. Consequently, the inequality

$$\epsilon - pv < 0$$

or

$$v > \epsilon/p. \quad (78.15)$$

must be fulfilled. This means that, if $\epsilon/p \neq 0$, excitation quanta can arise in the flowing helium, and an energy dissipation can occur only for a sufficiently large velocity of flow. For a velocity of flow which does not satisfy the inequality (78.15) no interaction accompanied by the appearance of thermal excitation quanta can arise between the wall and the helium.

From the form of the energy spectrum of helium II shown in fig. III.47 (§77) it is clear that the quantity ϵ/p for helium II is always different from zero. Thus, at the temperature of absolute zero liquid helium II moves along a solid wall without any interaction and energy dissipation, provided its velocity of motion does not exceed $v_0 = (\epsilon/p)_{\min}$, where $(\epsilon/p)_{\min}$ is the minimum value of the ratio ϵ/p . It is in this that the phenomenon of superfluidity consists.

For $T \neq 0$ all previous reasoning remains valid, and no new excitation quanta can arise in helium II when $v \neq v_0$. However, the already existing thermal excitation quanta can interact with a solid wall.

It turns out that in helium II at $T \neq 0$ two forms of motion are possible, superfluid motion and normal motion, which can occur in the same portion of the liquid simultaneously and independently of each other. The superfluid motion is without viscosity and is not accompanied by any transfer of thermal excitation energy. The normal flow proceeds in the same way as an ordinary flow of a liquid with a viscosity different from zero. With each of the forms of motion is associated the transfer of a part of the mass of the helium. Because

of this helium II can, in an obvious although not strictly accurate way, be considered as a mixture of two liquids: a superfluid liquid and a normal liquid. The motion of the superfluid liquid, which carries a part of the helium II at $T \neq 0$, is the same as the motion of the entire helium II at $T = 0$. However, at $T \neq 0$ a part of the mass of the helium is in the normal state, flows with friction and carries heat. The properties of the superfluid part of the liquid are shown in experiments on the flow of helium through fine capillary tubes. It flows through the finest capillary tube unimpeded. In experiments on the motion of bodies, for example in experiments on the oscillations of a disc immersed in a container of liquid helium, an interaction with the normal part of the helium is observed. In this case the motion of the disc is the same as in a normal liquid possessing viscosity. However, the mass of normal liquid turns out to depend on the temperature. As $T \rightarrow 0$ the mass of the normal part of helium II also decreases to zero.

One of the remarkable thermal properties of helium II is the so-called thermomechanical effect. This effect consists of the fact that, as helium flows out of a container through a very fine capillary tube, the temperature of the helium which remains in the container increases. On the contrary, as helium flows into a container the temperature in the container decreases.

The origin of the thermomechanical effect is understandable from the above. The superfluid part of the helium which moves through the fine capillary tube carries no thermal energy. When a certain amount of the superfluid helium flows out of the container the thermal energy content which existed before is distributed through the remaining amount and its temperature increases. As helium flows into the container the reverse phenomenon occurs: the thermal energy content which existed initially in the helium in the container is distributed over the entire amount of helium. The magnitude of the effect increases with decreasing temperature. This allows one to make use of the thermomechanical effect in helium to obtain very low temperatures.

§ 79. The electron gas at absolute zero

We shall now consider the behaviour of a Fermi system, an electron gas, at low temperatures. In expounding the quantum theory of metals in Part VI we shall show that in a certain approximation the behaviour of the whole set of electrons in metals can be described as the behaviour of an ideal degenerate Fermi gas. Hence the properties of the Fermi gas are of very great interest.

We shall first of all discuss the behaviour of an electron gas at absolute

zero. For this we write the Fermi distribution which, according to (72.11), has the form

$$dn = 2 \times 2\pi \left(\frac{2m}{h^2} \right)^{\frac{3}{2}} \frac{\epsilon^{\frac{1}{2}} d\epsilon}{\exp[(\epsilon - \mu)/kT] + 1} = \bar{n} \frac{d\gamma}{h^3} \quad (79.1)$$

The factor 2 is introduced in order to take into account the fact that to each energy level there correspond two states in which the electrons can have spins with opposite orientations.

We pass in (79.1) to the limit $T \rightarrow 0$. Then the Fermi distribution shown (for $T \neq 0$) in fig. III.48 assumes the form shown in fig. III.49 and is expressed by the formulae

$$\bar{n} = \begin{cases} 1 & \epsilon \leq \mu(T=0) = \epsilon_{\max} \\ 0 & \epsilon > \mu(T=0) = \epsilon_{\max} \end{cases} \quad (79.2)$$

where $\mu(T=0)$ denotes the chemical potential at absolute zero. We call this quantity the maximum energy at absolute zero, ϵ_{\max} . This result has a simple

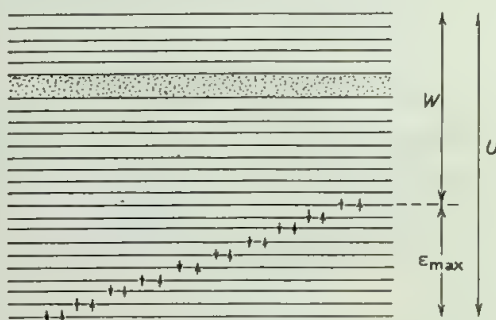


Fig. III.48

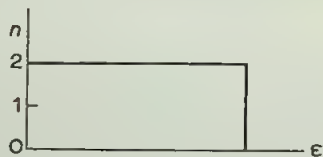


Fig. III.49

meaning: the energy levels of a system of a large number of electrons moving freely in a finite volume V , which is bounded by an impenetrable energy barrier (the walls), form an almost continuous spectrum. According to the exclusion principle, no more than two electrons with opposite spin orientations can simultaneously be in each energy level. Two electrons in a metal will occupy the lowest energy level with an energy equal to zero. The third and fourth electrons must be at the first excited level. Following electrons are distributed at higher energy levels, each successive electron pair occupying the corresponding level. If the total number of electrons in the system is equal to N , then the first $\frac{1}{2}N$ states with energies $0 \leq \epsilon \leq \epsilon_{\max}$ will be occupied at absolute zero. All other states with $\epsilon > \epsilon_{\max}$ will be free of electrons. This is schematically shown in fig. III.48.

Fig. III.49 shows the state distribution function for electrons at $T = 0$. The number of electrons in each occupied state is equal to two, and in unoccupied states it is equal to zero. It is obvious that, because of the exclusion principle, electrons must get into excited energy levels even at absolute zero.

Let us find the energy ϵ_{\max} of the highest of the occupied energy states of electrons at absolute zero. At each energy level there are two electrons, so that for the total number of occupied levels we have

$$4\pi \left(\frac{2m}{h^2} \right)^{\frac{3}{2}} V \int_0^{\epsilon_{\max}} \epsilon^{\frac{1}{2}} d\epsilon = N,$$

whence

$$N = \frac{8\pi}{3} V \left(\frac{2m\epsilon_{\max}}{h^2} \right)^{\frac{3}{2}} = \frac{8\pi V}{3h^3} p_{\max}^3$$

or

$$\epsilon_{\max} = \frac{h^2}{2m} \left(\frac{3N}{8\pi V} \right)^{\frac{2}{3}}. \quad (79.3)$$

The energy of all the electrons at absolute zero is obviously equal to

$$E_0 = 4\pi V \left(\frac{2m}{h^2} \right)^{\frac{3}{2}} \int_0^{\epsilon_{\max}} \epsilon^{\frac{3}{2}} d\epsilon = \frac{3}{5} N \epsilon_{\max}. \quad (79.4)$$

The mean energy of an individual electron in an electron gas at $T = 0$ amounts to $\frac{3}{5}$ of the maximum energy ϵ_{\max} .

The substitution of numerical values for the quantities contained in (79.3) gives, for example, for $N/V \sim 10^{19}$: $\epsilon_{\max} = 5$ eV or $\epsilon_{\max} = 6.0 \times 10^4$ degrees. One can also find the maximum velocity of electrons at absolute zero:

$$v_{\max} = P_{\max}/m = (2\epsilon_{\max}/m)^{\frac{1}{2}} = 1.39 \times 10^8 \text{ cm/sec} . \quad (79.3')$$

The velocities of electrons turn out to be very large even at absolute zero.

We see that the properties of an electron gas differ radically from those of a classical atomic gas.

The energy of an electron gas turns out to be proportional to the number of electrons N raised to the power $\frac{5}{3}$ and to the volume of the electron gas raised to the power $-\frac{2}{3}$. Electrons at absolute zero are not at rest, as is to be expected from classical concepts, but are moving with different velocities. The mean velocity of this motion is very large. In spite of this, the heat capacity of the electron gas at absolute zero turns out to be exactly equal to zero. Indeed,

$$C_V = \left(\frac{\partial E}{\partial T} \right)_{T=0} = 0 ,$$

since the energy of the gas does not depend on the temperature.

Electrons, as any other gas, exert a pressure on the walls of the container. For the gas pressure one can immediately write the expression (8.2):

$$p = \frac{2mN}{V} \int_0^\infty v_x^2 \rho(v_x) dv_x = \frac{Nm\overline{v^2}}{3V} , \quad (79.5)$$

since all directions of motion of electrons are equivalent and $\overline{v_x^2} = \frac{1}{3}\overline{v^2}$. For $\overline{v^2}$ it is necessary to substitute the expression

$$\overline{v^2} = 2E_0/mN , \quad (79.6)$$

where E_0/N is the mean energy per electron. Then we obtain

$$p = \frac{2}{3}E_0/V . \quad (79.7)$$

The same result can also be obtained by a thermodynamic method. The substitution of numerical values gives for $N/V \sim 10^{19}$: $p = 2 \times 10^5$ atm.

In the foregoing reasoning we have tacitly assumed that the electron gas is an ideal Fermi gas for which the interaction between particles can completely be disregarded. The validity of such an assumption may appear to be very questionable, particularly if it is taken into account that the density of the degenerate gas is very large and that it consists of charged particles for which the interaction decreases slowly with increasing distance.

The assumption that the gas is ideal is satisfied if the mean kinetic energy of an electron is very large in comparison with the mean energy of its interaction with other particles.

The mean kinetic energy of an electron is given by formula (79.4). The energy of interaction of two electrons is in order of magnitude equal to e^2/\bar{r} , where \bar{r} is the mean distance between them. If N/V is the number of electrons and ions per unit volume, then the mean distance \bar{r} is in order of magnitude equal to

$$\bar{r} \approx (V/N)^{1/3}.$$

The criterion of the smallness of the interaction energy can be written in the form

$$\frac{e^2}{(V/N)^{1/3}} \ll \epsilon_{\max}.$$

Since $\epsilon_{\max} \sim (N/V)^{2/3}$, then, upon performing simple transformations, we obtain

$$\frac{N}{V} \gg \left(\frac{e^2 m}{h^2} \right)^3.$$

For a large density of the electron gas the ratio of the interaction energy to the kinetic energy can turn out to be small.

Since the kinetic energy of the gas increases with increasing density more rapidly than the potential energy, we arrive at the following, at first sight paradoxical, result: for an electron gas to be considered as an ideal gas its density must be sufficiently great.

§80. The electron gas at low temperatures

We shall now consider the properties of an electron gas at temperatures which differ from zero but are still comparatively low. Namely, we shall

assume that the temperature of the metal is such that kT is considerably smaller than the maximum energy of electrons, ϵ_{\max} , at absolute zero. In this case the thermal excitation of the electron gas will be relatively negligible. This means that thermal excitation can bring electrons from energy levels occupied at $T=0$ only into nearby higher energy levels. It is clear, for example, that thermal excitation is insufficient to raise an electron from an energy level $\epsilon \ll \epsilon_{\max}$ to an energy level $\epsilon > \epsilon_{\max}$. It is only sufficient for the excitation of electrons which are at energy levels lying in a narrow interval of the order of magnitude of kT . Some of the electrons from these levels turn out to be raised to levels higher than the level $\epsilon = \epsilon_{\max}$ but above it by an amount which does not exceed kT .

Fig. III.50 shows schematically the thermal excitation at low temperature. Some of the levels lying below ϵ_{\max} turn out to be unoccupied for most of the time. Single electrons appear in levels lying above ϵ_{\max} . The state distribution function of electrons changes. At $T=0$ it is represented by the broken curve (fig. III.49) while at low temperatures above zero it assumes the form shown in fig. III.51. The distribution at $T=0$ is shown in fig. III.51 by a broken line. The fall of the curve for $\epsilon < \epsilon_{\max}$ means that the mean number of electrons in the corresponding levels turns out to be less than unity; electrons go into levels lying above ϵ_{\max} . The energy region in which the mean number of electrons at each level turns out to be smaller than unity but larger than zero is called the zone of diffuseness of the distribution function. From fig. III.51 and the above it is clear that the width of the zone of diffuseness is in order of magnitude equal to kT . The number of electrons appearing at levels lying above ϵ_{\max} is very small in comparison with the total number of electrons in the metal. Similarly the number of single electrons which are at various energy levels represents a small fraction of the total

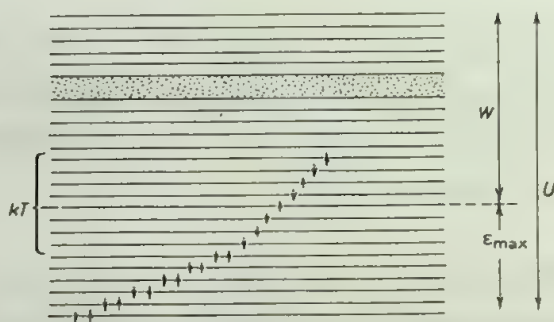


Fig. III.50

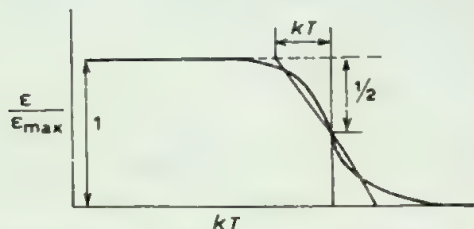


Fig. III.51

number of electrons. The condition of degeneracy $kT \ll \epsilon_{\max}$ is the same as that for a Fermi gas.

We now pass on to the discussion of the quantitative theory of an electron gas at a low temperature.

We confine ourselves to the calculation of two thermodynamic quantities, the chemical potential μ and the mean energy E , as functions of the temperature. The chemical potential of the electron gas is usually called the Fermi level or the Fermi surface. The origin of this terminology will be clear from what follows.

To determine the chemical potential we make use of the normalization condition. Namely, on the basis of (79.1) we can write

$$N = 2 \times 2\pi \left(\frac{2m}{h^3} \right)^{\frac{3}{2}} \int_0^{\infty} \frac{\epsilon^{\frac{1}{2}} d\epsilon}{\exp[(\epsilon - \mu)/kT] + 1}. \quad (80.1)$$

The mean energy of an electron gas is given by the formula

$$E = 4\pi \left(\frac{2m}{h^2} \right)^{\frac{3}{2}} V \int_0^{\infty} \frac{\epsilon^{\frac{3}{2}} d\epsilon}{\exp[(\epsilon - \mu)/kT] + 1}. \quad (80.2)$$

The integrals contained in formulae (80.1) and (80.2) are not taken in the general form. To calculate them at low temperatures we make use of the following method. Let

$$f = \left[\exp \left(\frac{\epsilon - \mu}{kT} \right) + 1 \right]^{-1} \quad (80.3)$$

and

$$I = \int_0^{\infty} f \epsilon^n d\epsilon, \quad n > 0.$$

Integrating by parts, we obtain

$$I = f \frac{\epsilon^{n+1}}{n+1} \Big|_0^{\infty} - \frac{1}{n+1} \int_0^{\infty} \epsilon^{n+1} \frac{\partial f}{\partial \epsilon} d\epsilon. \quad (80.4)$$

When the limits are substituted the first term obviously reduces to zero. The function $\partial f / \partial \epsilon$ is shown in fig. III.52. We see that it is an even function of its argument and has such a sharp maximum for $\epsilon = \mu$ that one can consider it as one of the representations of the δ -function. We introduce a new variable

$$x = \frac{\epsilon - \mu}{kT}.$$

Then

$$I = -\frac{1}{n+1} \int_{-\mu/kT}^{\infty} (\mu + kTx)^{n+1} \frac{\partial f}{\partial x} dx.$$

Since we are considering the low temperature region, the lower limit can be replaced by minus infinity:

$$I = -\frac{1}{n+1} \int_{-\infty}^{\infty} (\mu + kTx)^{n+1} \frac{\partial f}{\partial x} dx.$$

In that region of variation of x in which $\partial f / \partial x$ has a value differing from zero, i.e. for $\epsilon \approx \mu$, the value of x is very small. Hence the first factor in the inte-

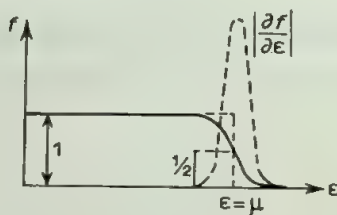


Fig. III.52

grand can be expanded in a series and one can restrict oneself to the first terms of the expansion. For large x , when this cannot be done, the integrand reduces to zero owing to the factor $\partial f/\partial x$ which is negligibly small everywhere except for $x = 0$. Thus,

$$I \approx -\frac{1}{n+1} \int_{-\infty}^{\infty} \mu^{n+1} \left[1 + (n+1) \frac{kTx}{\mu} + \frac{(n+1)n}{2} \left(\frac{kT}{\mu} \right)^2 x^2 + \dots \right] \frac{\partial f}{\partial x} dx. \quad (80.5)$$

Since $\partial f/\partial x$ is an even function, we have

$$\int_{-\infty}^{\infty} x \frac{\partial f}{\partial x} dx = 0, \quad \int_{-\infty}^{\infty} x^2 \frac{\partial f}{\partial x} dx = 2 \int_0^{\infty} x^2 \frac{\partial f}{\partial x} dx = 0.$$

Hence

$$I = -\frac{1}{n+1} \mu^{n+1} \left[\int_{-\infty}^{\infty} \frac{\partial f}{\partial x} dx + \frac{(n+1)n}{2} \left(\frac{kT}{\mu} \right)^2 \int_{-\infty}^{\infty} x^2 \frac{\partial f}{\partial x} dx + \dots \right]. \quad (80.6)$$

It is easily seen that

$$\int_{-\infty}^{\infty} \frac{\partial f}{\partial x} dx = f \Big|_{-\infty}^{\infty} = -1.$$

The second integral in (80.6) is calculated in Appendix IV. It is equal to $-\frac{1}{3}\pi^2$. Finally we have

$$I = \frac{\mu^{n+1}}{n+1} \left[1 + \frac{(n+1)n}{6} \pi^2 \left(\frac{kT}{\mu} \right)^2 \right]. \quad (80.7)$$

The discarded terms are proportional to higher powers of the ratio kT/μ .

Coming back to the normalization condition in which $n = \frac{1}{2}$, we have

$$\begin{aligned} N &= 4\pi \left(\frac{2m}{h^2} \right)^{\frac{3}{2}} V \int_0^{\infty} \frac{\epsilon^{\frac{1}{2}} d\epsilon}{\exp[(\epsilon - \mu)/kT] + 1} \approx \\ &\approx 4\pi \left(\frac{2m}{h^2} \right)^{\frac{3}{2}} \frac{2}{3} V \mu^{\frac{3}{2}} \left[1 + \frac{\pi^2}{8} \left(\frac{kT}{\mu} \right)^2 \right]. \end{aligned} \quad (80.8)$$

For $T = 0$ the second term in (80.8) reduces to zero, and

$$N = \frac{8\pi}{3} \left(\frac{2m}{h^2} \right)^{\frac{3}{2}} \mu_{T=0}^{\frac{3}{2}} V. \quad (80.9)$$

Comparing (80.9) with (79.3), we find that

$$\mu_{T=0} = \mu_0 = \epsilon_{\max}. \quad (80.10)$$

The chemical potential of the electron gas at absolute zero turns out to be equal to the maximum energy of an electron at absolute zero.

For temperatures close to absolute zero, (80.8) can be solved for μ by a method of successive approximations. In the small second term of (80.8) μ can be replaced by $\mu_0 = \epsilon_{\max}$. Then we have

$$\mu = \mu_0 \left[1 - \frac{\pi^2}{12} \frac{(kT)^2}{\mu_0^2} \right]. \quad (80.11)$$

Analogously, for the mean energy of the electron gas ($n = \frac{3}{2}$) we find

$$E \approx 4\pi \left(\frac{2m}{h^2} \right)^{\frac{3}{2}} V \left[\frac{2}{5} \mu^{\frac{5}{2}} + \frac{1}{3} \pi^2 \mu^{\frac{1}{2}} (kT)^2 \right]. \quad (80.12)$$

Substituting the value of μ from (80.11) into (80.12) and disregarding the higher powers of the ratio kT/μ , we obtain

$$E = \frac{3}{5} N \epsilon_{\max} \left[1 + \frac{5\pi^2}{12} \left(\frac{kT}{\epsilon_{\max}} \right)^2 \right]. \quad (80.13)$$

From (80.9) or (80.10) it can be seen that the conditions of degeneracy $kT \ll \epsilon_{\max}$ and (72.16) are completely identical. Hence the electron gas is degenerate at temperatures $T < \epsilon_{\max}/k$. We have pointed out that for $N/V \sim 10^{19}$ the maximum energy ϵ_{\max} expressed in units of temperature is equal to 6×10^4 .

From (80.13) one can find the heat capacity of the electron gas:

$$C_V = \frac{Nk\pi^2}{2} \frac{kT}{\epsilon_{\max}}. \quad (80.14)$$

The heat capacity of the electron gas turns out to be a linear function of the

temperature, and reduces to zero at $T = 0$. The factor of proportionality contains only known quantities: universal constants and the density of the electron gas N/V . For simple metals, for example, copper and silver, whose atoms have one weakly bound valence electron, the number of free electrons per atom can be considered to be equal to unity. Then, for example, for copper the theoretical expression for the electron heat capacity has the form

$$C_V = 0.9 \times 10^{-4} NkT. \quad (80.15)$$

Thus, the heat capacity of the electron gas, in agreement with experimental data, is very small and at normal temperature represents an immeasurably small fraction of the heat capacity of the crystal lattice.

The heat capacity of a lattice at temperatures which are lower than the characteristic temperature rapidly decreases and tends to zero as the cube of the temperature when $T \rightarrow 0$. The heat capacity of an electron gas also tends to zero as $T \rightarrow 0$, but much more slowly, as the first power of the temperature. The ratio of the heat capacity of the electron gas to that of the lattice is equal to

$$\frac{C_V^{\text{el}}}{C_V^{\text{lat}}} = \frac{5}{24\pi^2} \frac{kT}{\epsilon_{\text{max}}} \left(\frac{\theta}{T} \right)^3.$$

For copper $\epsilon_{\text{max}} = 5$ eV and the characteristic temperature is $\theta = 335$ K, so that

$$\frac{C_V^{\text{el}}}{C_V^{\text{lat}}} = 0.02 \frac{T}{7 \times 10^4} \left(\frac{335}{T} \right)^3.$$

The ratio between the heat capacities is of the order of unity at $T = 3.3$ K. At still lower temperatures the heat capacity of the electrons turns out to be larger than the heat capacity of the lattice.

Accurate measurements of the heat capacity confirmed the validity of the theoretical formulae. Thus for copper, for which the measurements are particularly accurate, the heat capacity at low temperatures turned out to be made up of two terms, one of which agreed very accurately with the theoretical formula for the heat capacity of the lattice. The second term agrees with an accuracy to within 2% with the theoretical formula (80.14). At temperatures which are lower than 3 K the electron heat capacity is larger than the heat capacity of the lattice and is superposed on the theoretical curve.

In conclusion we shall calculate the number of electrons in the zone of diffuseness of the Fermi distribution, which it is convenient to do for an obvious interpretation of a number of formulae in what follows.

The number of electrons in the zone of diffuseness, or the number of unpaired electrons, will also be called by us the number of effective electrons n_{eff} . It is these electrons which can change their state under the effect of external actions. Hence the heat capacity of the electron gas, its electric conductivity and so on depend on effective electrons.

The value of n_{eff} can be found from the following. The probability for an electron to be found in a given state with an energy ϵ is proportional to the value of the distribution function f . The probability that this state is not occupied is equal to $1 - f$. Since only electrons with antiparallel spins can be in one state, the product of the probabilities $f(1-f)$ represents the probability that in the state with energy ϵ there is one electron, while the second electron with the antiparallel spin is not in it. In other words, $f(1-f)$ represents the probability that the state with energy ϵ will be occupied by only one electron. The total number of such states or, what is the same, the total number of unpaired electrons is equal to the integral of the product of $f(1-f)$ and the number of states $d\Omega$ with given energy, the integral being taken over all values of the energy:

$$n_{\text{eff}} = 2 \int_0^{\infty} f(1-f) d\Omega = \frac{4\pi(2m)^{\frac{3}{2}}}{h^3} \int_0^{\infty} f(1-f) \epsilon^{\frac{1}{2}} d\epsilon. \quad (80.16)$$

In a strongly degenerate gas $\mu \gg kT$, so that

$$f = \left[\exp \frac{\epsilon - \mu}{kT} + 1 \right]^{-1} \approx 1, \quad 1 - f \approx \exp \frac{\epsilon - \mu}{kT}.$$

For $\epsilon - \mu \gg kT$ the integrand decreases exponentially. Hence, instead of integrating up to infinitely large values of the energy, one can confine oneself to an integration up to a value $\epsilon \approx \mu$. Then

$$n_{\text{eff}} \approx \frac{4\pi(2m)^{\frac{3}{2}}}{h^3} \int_0^{\mu} \left(\exp \frac{\epsilon - \mu}{kT} \right) \epsilon^{\frac{1}{2}} d\epsilon.$$

Since the exponential factor rapidly decreases, one can bring the slowly varying factor $\epsilon^{\frac{1}{2}}$ out of the sign of the integral, taking its value at the upper limit. This gives

$$\int_0^{\mu} \left(\exp \frac{\epsilon - \mu}{kT} \right) \epsilon^{\frac{1}{2}} d\epsilon \approx \mu^{\frac{1}{2}} \int_0^{\mu} \exp \frac{\epsilon - \mu}{kT} d\epsilon = kT\mu^{\frac{1}{2}} \left(1 - \exp -\frac{\mu}{kT} \right) \approx$$

$$\approx kT\mu^{\frac{1}{2}} \approx kT\mu_0^{\frac{1}{2}}$$

since $\mu \gg kT$, whence

$$n_{\text{eff}} \approx \frac{4\pi(2m)^{\frac{3}{2}} kT\mu_0^{\frac{1}{2}}}{h^3}. \quad (80.17)$$

By means of formulae (79.3) and (80.10) n_{eff} can be written in the form

$$n_{\text{eff}} = \frac{3NkT}{2V\epsilon_{\text{max}}}. \quad (80.18)$$

Thus, n_{eff} represents (with an accuracy to within the numerical coefficient) a small fraction of the total number of electrons per cm^3 which is equal approximately to kT/ϵ_{max} .

By means of n_{eff} the heat capacity of the electron gas in order of magnitude can be written in the form

$$C_V^{\text{el}} \sim \frac{3}{2} k n_{\text{eff}}, \quad (80.19)$$

where $\frac{3}{2}k$ is the classical value of the heat capacity.

Formula (80.19) shows that the properties of an electron gas can be approximately interpreted in the following way. In the gas there are n_{eff} effective particles which can change their state and absorb energy supplied from without. Every one of these particles possesses ordinary classical properties and, in particular, to every one of these there corresponds the usual value of the heat capacity.

PART IV

ELECTROMAGNETIC PROCESSES
IN MATTER

Electromagnetic Fields in Matter

§ 1. The derivation of the basic field equations

We have already considered electromagnetic processes occurring in a vacuum. The motion of electric charges in a vacuum and the electromagnetic field accompanying them were considered. We now pass on to the study of electromagnetic phenomena in matter (i.e. in a medium). The theory of electromagnetic processes in matter is often called macroscopic electrodynamics.

The character of electromagnetic processes in matter depends essentially on the properties of the matter. For example, the mechanisms of the passage of current through metal conductors and gaseous plasmas are essentially different in character and are accompanied by different phenomena; magnetic phenomena in ferromagnetics differ strongly from those in diamagnetics and paramagnetics, and so on. Nevertheless, it turns out to be possible to construct a phenomenological theory of electromagnetic phenomena in matter on the basis of certain very general assumptions. It is necessary, first of all, to find the general equations of the electromagnetic field in matter. For this, as we shall see below, it will be necessary to make certain very general assumptions about the concrete properties of the medium in which one or other electromagnetic process takes place.

We have seen that the basic equations of the theory of the electromagnetic field, the Maxwell–Lorentz equations, contain quantities referring to a given point and a given instant of time. We shall write these equations in a slightly different notation, replacing \mathbf{E} by \mathbf{e} and \mathbf{H} by \mathbf{h} . We write the Maxwell–Lorentz equations as well as the charge conservation law in this new notation:

$$\nabla \times \mathbf{e} = -\frac{1}{c} \frac{\partial \mathbf{h}}{\partial t}, \quad (1.1)$$

$$\nabla \cdot \mathbf{h} = 0, \quad (1.2)$$

$$\nabla \times \mathbf{h} = \frac{4\pi}{c} \rho \mathbf{v} + \frac{1}{c} \frac{\partial \mathbf{e}}{\partial t}, \quad (1.3)$$

$$\nabla \cdot \mathbf{e} = 4\pi\rho, \quad (1.4)$$

$$\nabla \cdot \rho \mathbf{v} + \frac{\partial \rho}{\partial t} = 0. \quad (1.5)$$

In matter, a medium consisting of atoms or molecules, the Maxwell–Lorentz equations characterizing the field at a given point and a given instant of time lose their meaning. Indeed, in matter all quantities, including also the electromagnetic field, vary very rapidly from point to point and at a given point also vary rapidly in time. For example, the electric field strength has a relatively small value outside any given atom, becomes very large inside the atom, and again decreases outside it. The increase of the field by a factor of several million and its subsequent decrease take place on a scale of the order of magnitude of atomic dimensions. A change in the field with time at a fixed point occurs, for example, because of the thermal motion of the atom in small fractions of a second. Hence, as in other macroscopic processes taking place in matter, only the mean values of the corresponding quantities are of interest and significant (see Part III).

Let us average the Maxwell–Lorentz equations over a volume v_0 large compared with atomic dimensions but macroscopically small and a time interval τ , introducing the mean values according to the formula

$$\bar{f} = \frac{1}{2\tau} \int_{-\tau}^{\tau} \frac{1}{v_0} \int f dV dt. \quad (1.6)$$

We then have

$$\nabla \times \bar{\mathbf{e}} = -\frac{1}{c} \frac{\partial \bar{\mathbf{h}}}{\partial t}, \quad (1.7)$$

$$\nabla \cdot \bar{\mathbf{h}} = 0, \quad (1.8)$$

$$\nabla \times \bar{\mathbf{h}} = \frac{4\pi}{c} \bar{\rho} \bar{\mathbf{v}} + \frac{1}{c} \frac{\partial \bar{\mathbf{e}}}{\partial t}, \quad (1.9)$$

$$\nabla \cdot \bar{\mathbf{e}} = 4\pi \bar{\rho}, \quad (1.10)$$

$$\nabla \cdot \bar{\rho} \bar{\mathbf{v}} + \frac{\partial \bar{\rho}}{\partial t} = 0. \quad (1.11)$$

We introduce the following notation:

$$\bar{\mathbf{e}} = \mathbf{E}, \quad (1.12)$$

$$\bar{\mathbf{h}} = \mathbf{B}. \quad (1.13)$$

The field strength \mathbf{E} denotes the mean value of the electric field strength in the medium, and the magnetic induction \mathbf{B} the mean value of the magnetic field strength in the medium (this name for the mean magnetic field is associated solely with historical tradition). Then eqs. (1.7) – (1.11) assume the form

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad (1.14)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (1.15)$$

$$\nabla \times \mathbf{B} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \bar{\rho} \bar{\mathbf{v}}, \quad (1.16)$$

$$\nabla \cdot \mathbf{E} = 4\pi \bar{\rho}, \quad (1.17)$$

$$\nabla \cdot \bar{\rho} \bar{\mathbf{v}} + \frac{\partial \bar{\rho}}{\partial t} = 0. \quad (1.18)$$

Further transformations of the equations are associated with the calculation of the mean values ρ and ρv . To find these mean values it is necessary to introduce certain assumptions about the structure of matter.

In the theory of the electromagnetic field in matter, the latter is considered as a continuous medium whose properties are described by means of a number of formal macroscopic characteristics, such as the dielectric constant, conductivity and so on. Some of these formal characteristics can be found by applying the methods of statistical physics and are connected with the molecular structure of matter.

However, we shall, from the very beginning, have to divide all matter into two groups; conductors and dielectrics. By conductors we shall understand bodies in which, under the action of an external static field, a displacement of charges occurs over the volume and an electric current arises corresponding to this motion of charges. In dielectrics an external field does not give rise to the motion of charges, although it can cause their displacement to new equilibrium positions.

It is clear from these definitions that the division of bodies into conductors and dielectrics is somewhat arbitrary. In practice, the external electric field gives rise to a finite, although very small, current in dielectrics. On the other hand, in some conductors also the current can be small. Semiconductors play a most important role in physics and technology. Under certain conditions the electric current in semiconductors is relatively large and approaches that in conductors. At other times the current in semiconductors is as small as in the best dielectrics.

Nevertheless, the schematic separation of all bodies into conductors and dielectrics appears to be a sufficiently good approximation on the basis of which it is possible to construct a phenomenological theory of electromagnetic phenomena in continuous media.

§2. The polarization of a medium in an electric field

In calculating ρ one has to distinguish between the cases of a body which is as a whole electrically neutral and one having a charge different from zero.

Let us consider the first case. If an electrically neutral body is placed in an external electric field, a displacement with respect to each other of the positive and negative charges occurs in the atoms and molecules. The body, remaining electrically neutral, acquires a dipole moment. We shall characterize the body by the total dipole moment of all its particles in unit volume. The total dipole moment per unit volume will be called the polariza-

tion vector or, briefly, polarization \mathbf{P} . The dipole moment acquired by the body is by definition equal to

$$\mathbf{d} = \int \mathbf{P} dV = \int \rho_{\text{bound}} \mathbf{r} dV. \quad (2.1)$$

In view of the arbitrariness of the volume of integration

$$\mathbf{P} = \rho_{\text{bound}} \mathbf{r}. \quad (2.1')$$

If the body is uniformly polarized, i.e. if \mathbf{P} is the same at all points of the body, then $\mathbf{P} = \mathbf{d}/V$. The appearance of polarization in a body can, under certain conditions which will be clear from what follows, be accompanied by the appearance of a mean charge density ρ_{bound} . The subscript "bound" means that the mean charge density is due to the displacement of charges which are bound in the atoms of the body. The appearance of the bound, or induced, charge is called electrostatic induction or polarization.

To find ρ_{bound} we shall make use of the definition (2.1) and shall transform the integral $\int \mathbf{P} dV$ into the form $\int f(\mathbf{P}) \mathbf{r} dV$, where $f(\mathbf{P})$ is a function of the polarization. Then, comparing $\int f(\mathbf{P}) \mathbf{r} dV$ with $\int \rho_{\text{bound}} \mathbf{r} dV$, one can find the value of ρ_{bound} . Making use of formula (1.53) and assuming in it that $\mathbf{a} = \mathbf{P}$, $\mathbf{b} = \mathbf{r}$, we find:

$$\int (\mathbf{P} \cdot \mathbf{n}) \mathbf{r} dS = \int \mathbf{r} (\nabla \cdot \mathbf{P}) dV + \int (\mathbf{P} \cdot \nabla) \mathbf{r} dV = \int \mathbf{r} (\nabla \cdot \mathbf{P}) dV + \int \mathbf{P} dV.$$

Choosing the surface of integration to be outside the volume occupied by the body, we have

$$(\mathbf{P} \cdot \mathbf{n}) \mathbf{r} dS = 0,$$

since outside the body the polarization is equal to zero. Hence

$$\int \mathbf{P} dV = - \int \mathbf{r} (\nabla \cdot \mathbf{P}) dV. \quad (2.2)$$

Comparing (2.2) with (2.1), we obtain

$$\rho_{\text{bound}} = - \nabla \cdot \mathbf{P}. \quad (2.3)$$

Thus, if the polarization of the body is not uniform and the vector \mathbf{P} varies from point to point in such a way that $\nabla \cdot \mathbf{P} \neq 0$, a charge density ρ_{bound} defined by formula (2.3) arises in the body.

Formula (2.3) can be given an obvious meaning. Let there be inside the body two imaginary planes S_1 and S_2 . In an external field the charges which are bound in the atoms confined in the space between the planes are displaced. If the polarization is not uniform in space, then, for example, the surface S_2 will be crossed by a larger number of charges going out of the volume than the surface S_1 . As a result, the number of charges going out of the volume confined between the planes is larger than that of charges coming in from the neighbouring volume. A charge density ρ_{bound} arises in the space between S_2 and S_1 .

In particular, when the surface is the boundary surface of the body, then the charges arising from polarization form a surface charge. Its density ω_{bound} is determined in the following simple calculation. We integrate (2.3) with respect to the volume confined between the two surfaces. We choose these surfaces in such a way that one of them is outside the body and the other inside it and that they are at a very small distance h from the external surface of the body.

Then we have

$$\begin{aligned} \int (\nabla \cdot \mathbf{P}) dV &= \oint \mathbf{P} \cdot d\mathbf{S} = \oint_{\substack{\text{outside} \\ \text{the body}}} P_n dS - \oint_{\substack{\text{inside} \\ \text{the body}}} P_n dS = \\ &= - \oint_{\substack{\text{inside} \\ \text{the body}}} P_n dS = - \int \rho_{\text{bound}} dV. \end{aligned} \quad (2.4)$$

Passing to the limit $h \rightarrow 0$, we can write

$$\begin{aligned} \lim_{h \rightarrow 0} \int \rho_{\text{bound}} dV &= \oint_{S'} \omega_S dS, \\ \oint_{\substack{\text{inside} \\ \text{the body}}} P_n dS &\rightarrow \int_{S'} P_n dS, \end{aligned}$$

where S denotes the surface of the body and the value of P_n is taken at the surface. In view of the arbitrary character of the surface of the body, it follows from the equality

$$\oint_S P_n dS = \int \omega_S dS,$$

that

$$\omega_S = P_n. \quad (2.5)$$

In this case it is clear that the total charge of an electrically neutral body placed in an external field remains equal to zero. Choosing the surface of integration in (2.4) to be outside the volume of the body, one can write

$$\int \rho_{\text{bound}} dV = - \int (\mathbf{V} \cdot \mathbf{P}) dV = - \int P_n dS = 0.$$

Thus, the total bound charge density arising within the body is equal to the total charge induced at its surface.

Up to now we have assumed the body to be electrically neutral. If the total free charge of the body is different from zero and is distributed in it with a volume density ρ , then the mean charge density is

$$\bar{\rho} = \rho_{\text{bound}} + \rho$$

and

$$\int \bar{\rho} dV = \int (\rho_{\text{bound}} + \rho) dV = \int \rho dV = e. \quad (2.6)$$

The charge of the body characterized by the density ρ is not bound to the atoms of the substance and is not induced by an external field. We shall see below that in a constant field the free charge ρ can exist only in dielectrics. In conductors free charges are mobile and move until they come to the surface of the conductor, forming a surface charge.

§3. The mean current density and mean charge density in a medium

The calculation of the mean current density $\overline{\rho \mathbf{v}}$ in a medium is a somewhat more complex problem. This calculation can be carried out either on the basis of certain model representations or, more formally, by proceeding from the general notions of the electromagnetic properties of a medium.

We choose the second method here, since the atomic and molecular models which are usually employed in expounding macroscopic electrodynamics are not only far from reality but also contain of necessity obviously incorrect assumptions. As will be explained below, quantum effects play a fundamental role in the magnetic properties of atoms. Hence the consideration of the electric and, in particular, the magnetic properties of atomic systems cannot be carried out on the basis of classical models. In a formal treatment we can

restrict ourselves to one physical assumption only: if a body is placed in an external electromagnetic field, then the mean field within the body is small in comparison with the intra-atomic fields. In other words, we shall assume that mean fields inside the body are weak. In addition, we shall confine ourselves to cases of homogeneous and isotropic bodies.

The mean current density, $\bar{\rho}\mathbf{v}$, at each point of the body is a function of the strengths of the electric and magnetic fields. In addition, if the fields vary in space and time, the mean current density will depend on the rate of change in time of the vectors \mathbf{E} and \mathbf{B} and on the spatial derivatives of these quantities, i.e.

$$\bar{\rho}\mathbf{v} = f\left(\mathbf{E}, \mathbf{B}, \frac{\partial \mathbf{E}}{\partial t}, \frac{\partial \mathbf{B}}{\partial t}, \frac{\partial E_i}{\partial x_k}, \frac{\partial H_i}{\partial x_k}\right). \quad (3.1)$$

Since the fields are weak, one can expand the function f in a series in powers of the variables and confine oneself to the first powers of the expansion*. This expansion is carried out in powers of a small ratio of the type $|\mathbf{E}|/|\mathbf{E}_{\text{int.at.}}|$, where $E_{\text{int.at.}}$ is the strength of the intra-atomic field.

Expanding f in a series in powers of the variables, we have to take into account that $\rho\mathbf{v}$ is a polar vector. Hence all terms of the series must also represent polar vectors. They can be neither scalars nor axial vectors.

We recall (see §6 of Part I) that the electric field strength \mathbf{e} and, consequently, also the mean electric field strength \mathbf{E} in a medium are polar vectors. On the contrary, the magnetic field strength \mathbf{h} and the mean magnetic field \mathbf{B} are axial vectors or pseudovectors. Hence the vector \mathbf{E} but not the vector \mathbf{B} can appear in the required expansion.

The derivative of the vector \mathbf{E} with respect to time, $\partial\mathbf{E}/\partial t$, is a polar vector and must be the first term of the expansion. The spatial derivatives $\partial E_i/\partial x_k$ can be grouped in the form of two combinations of the derivatives, $\nabla \times \mathbf{E}$ and $\nabla \cdot \mathbf{E}$. The first combination of the derivatives, $\nabla \times \mathbf{E}$, forms an axial vector, while the second forms a scalar. These two quantities cannot by themselves occur in the expansion. The first derivatives of \mathbf{B} are: $\partial\mathbf{B}/\partial t$, $\nabla \times \mathbf{B}$ and $\nabla \cdot \mathbf{B}$. The quantity $\partial\mathbf{B}/\partial t$ is an axial vector, $\nabla \cdot \mathbf{B} = 0$ by virtue of (1.15), while $\nabla \times \mathbf{B}$ is a polar vector. Hence the vector $\nabla \times \mathbf{B}$ must be retained in the expansion f . Thus, to the first order of small quantities, there are only three quantities which are polar vectors: \mathbf{E} , $\partial\mathbf{E}/\partial t$ and $\nabla \times \mathbf{B}$. Of course, in the next approximation one can form a number of polar vectors

* See I.E. Tamm, *Osnovy teorii elektrichestva (Introduction to the theory of electricity)* (Gostekhizdat, Moscow, 1954) p. 428.

from the scalars and axial vectors, but we shall not be interested in the expansion terms of the second order.

In expanding in powers of the ratio $|\mathbf{E}|/|\mathbf{E}_{\text{int.at.}}|$ one has, in general, to take into account the anisotropy of the body, because $\mathbf{E}_{\text{int.at.}}$ varies in different directions according to different laws. In isotropic bodies (such are gases, liquids and, to a certain extent, polycrystalline bodies) the ratio $|\mathbf{E}|/|\mathbf{E}_{\text{int.at.}}|$ has the same value in all directions. Then, with an accuracy to terms of the first order of small quantities,

$$\bar{\rho}\mathbf{v} = \sigma\mathbf{E} + \chi \frac{\partial \mathbf{E}}{\partial t} + \alpha c (\nabla \times \mathbf{B}), \quad (3.2)$$

where σ , χ and α are scalars depending on the properties of the medium. The reason why the factor c (the velocity of light) is introduced into the last term will be clear from what follows.

The zeroth order term (the term containing no field) is absent in expansion (3.2), since a mean current equal to zero must correspond to the absence of a field.

Before introducing this expression for $\rho\mathbf{v}$ into Maxwell's equations, we shall discuss the physical meaning of individual terms of the expansion.

The first term of the expansion, $\sigma\mathbf{E}$, has a very simple and obvious meaning. It shows that in the presence of an electric field a current arises in the medium whose mean density is proportional to the mean field strength \mathbf{E} . The quantity σ is called the conductivity or electrical conductivity of the body.

We shall make use of the notation

$$\mathbf{j} = \sigma\mathbf{E}. \quad (3.3)$$

The vector \mathbf{j} represents the mean charge passing per second through 1 cm^2 of surface inside the body perpendicular to \mathbf{j} . The vector \mathbf{j} is called the conduction current density. The relation (3.3) connects the current density at a given point of the body with the field strength at that point. We shall call the relation (3.3) Ohm's law in differential form. Below, in § 16, we shall establish the connection of (3.3) with the usual formulation of Ohm's law.

The properties of bodies are characterized by the value of the conductivity σ . For ideal dielectrics it should be assumed, in accordance with their definition, that $\sigma = 0$. For real dielectrics $\sigma \neq 0$, but is very small in comparison with its value in semiconductors and particularly in metals.

The problem of conductivity in bodies of different natures will be

considered below. Meanwhile we note only, that the conduction current is due to free charges: free electrons in metals, ions and electrons in gases, ions in liquids. They are called free because they are not bound to any atom and, under the action of a field, can move over the entire volume of the body.

We shall denote the mean density of free charges in a body by ρ . It is obviously connected with the current density \mathbf{j} by the equation of continuity

$$\nabla \cdot \mathbf{j} + \partial \rho / \partial t = 0. \quad (3.4)$$

It is here assumed that the free charges existing in a body do not become bound and do not arise from bound charges.

In order to explain the physical meaning of the second and third terms of the expansion (3.2) we transform (3.2) in such a way that the vectors \mathbf{E} and \mathbf{B} are separated from each other. We take, for example, the divergence of both sides of eq. (3.2). We then have, obviously,

$$\nabla \cdot (\bar{\rho} \mathbf{v}) = \nabla \cdot \mathbf{j} + \chi \frac{\partial}{\partial t} (\nabla \cdot \mathbf{E}).$$

By virtue of the equations of continuity for the total current $\rho \mathbf{v}$ and the free charge current \mathbf{j} , given by formulae (1.18) and (3.4),

$$-\frac{\partial}{\partial t} (\bar{\rho} - \rho) = \chi \frac{\partial}{\partial t} \nabla \cdot \mathbf{E}. \quad (3.5)$$

The difference $\bar{\rho} - \rho$ is obviously equal to the mean density of bound charges ρ_{bound} . Thus, for a homogeneous medium we have

$$\frac{\partial \rho_{\text{bound}}}{\partial t} = -\frac{\partial}{\partial t} \nabla \cdot (\chi \mathbf{E}).$$

Integrating, one can write

$$\rho_{\text{bound}} = -\nabla \cdot (\chi \mathbf{E}). \quad (3.6)$$

The integration constant may be set equal to zero, since in the absence of a field ρ_{bound} must be equal to zero.

Comparing (3.6) with (2.3), we verify, first of all, the formula

$$\mathbf{P} = \chi \mathbf{E}, \quad (3.7)$$

which relates the polarization in the body to the field strength \mathbf{E} . The polarization \mathbf{P} turns out to be proportional to the field \mathbf{E} . The factor of proportionality χ is called the polarization coefficient or dielectric susceptibility of the body. The value of the dielectric susceptibility for bodies consisting of simple molecules will be calculated in §12 and §13. There it will be shown that χ is an essentially positive quantity, so that the vector \mathbf{P} is always directed in the same direction as the vector \mathbf{E} .

By virtue of (3.7) the second term of (3.2) can be written in the form

$$\chi \frac{\partial \mathbf{E}}{\partial t} = \frac{\partial \mathbf{P}}{\partial t}. \quad (3.8)$$

Formula (3.8) shows that the variation of the polarization vector in time is equivalent to the appearance of a certain current. This current is called the polarization current, and its density is

$$\mathbf{j}_{\text{pol}} = \partial \mathbf{P} / \partial t, \quad (3.9)$$

i.e. is equal to the rate of change of the dipole moment per unit volume at a given point of the body.

Indeed, from the definition (2.1) of the vector \mathbf{P} it follows that

$$\int \frac{\partial \mathbf{P}}{\partial t} dV = \frac{\partial}{\partial t} \int \rho_{\text{bound}} \mathbf{r} dV = \int \mathbf{r} \frac{\partial \rho_{\text{bound}}}{\partial t} dV. \quad (3.10)$$

Here we have changed the order of differentiation and integration. The quantity \mathbf{r} represents the integration variable and does not depend on t .

Making use of (1.5) and applying the relation (2.2) to the vector $\rho_{\text{bound}} \mathbf{v}$, we find

$$\int \frac{\partial \mathbf{P}}{\partial t} dV = - \int \mathbf{r} [\nabla \cdot (\rho_{\text{bound}} \mathbf{v})] dV = \int \rho_{\text{bound}} \mathbf{v} dV. \quad (3.11)$$

The meaning of the density of the current of bound charges is very simple. The variation in time of the polarization of a given volume means that bound charges are going out of it (or coming into it). It is clear, however, that the displacement of bound charges, from the point of view of the electricity which they transport, is equivalent to the motion of free charges.

We now go on to the explanation of the meaning of the last term in (3.2). For this we form the vector product of formula (3.2) and \mathbf{r} and integrate the result with respect to the volume of the body:

$$\int \mathbf{r} \times (\overline{\rho \mathbf{v}}) dV = \int \mathbf{r} \times \mathbf{j} dV + \int \mathbf{r} \times \frac{\partial \mathbf{P}}{\partial t} dV + \alpha c \int \mathbf{r} \times (\nabla \times \mathbf{B}) dV.$$

By virtue of (3.10):

$$\begin{aligned} \int \mathbf{r} \times \frac{\partial \mathbf{P}}{\partial t} dV &= \int \mathbf{r} \times \left(\frac{\partial}{\partial t} (\rho_{\text{bound}} \mathbf{r}) \right) dV = \\ &= \frac{\partial}{\partial t} \int \mathbf{r} \times (\rho_{\text{bound}} \mathbf{r}) dV = \frac{\partial}{\partial t} \int \rho_{\text{bound}} (\mathbf{r} \times \mathbf{r}) dV = 0. \end{aligned}$$

Consequently,

$$\alpha c \int \mathbf{r} \times (\nabla \times \mathbf{B}) dV = \int \mathbf{r} \times (\overline{\rho \mathbf{v}} - \mathbf{j}) dV = \int \mathbf{r} \times (\rho_{\text{bound}} \mathbf{v}) dV, \quad (3.12)$$

where $\rho_{\text{bound}} \mathbf{v} = \overline{\rho \mathbf{v}} - \mathbf{j}$ is the current density carried by bound charges.

We introduce the notation

$$\mathbf{M} = \alpha \mathbf{B}. \quad (3.13)$$

Then (3.12) can be written in the form

$$\int \mathbf{r} \times (\nabla \times \mathbf{M}) dV = \frac{1}{c} \int \mathbf{r} \times (\rho_{\text{bound}} \mathbf{v}) dV. \quad (3.14)$$

By virtue of formula (1.54), in which $\mathbf{a} = \mathbf{M}$, $\mathbf{b} = \mathbf{r}$,

$$\int \mathbf{r} \times (\nabla \times \mathbf{M}) dV = - \int (\mathbf{M} \times \nabla) \times \mathbf{r} dV - \oint (\mathbf{n} \times \mathbf{M}) \times \mathbf{r} dV.$$

If the integration surface is outside the volume occupied by the body, then at this surface $\mathbf{M} = 0$ and, thus,

$$\int \mathbf{r} \times (\nabla \times \mathbf{M}) dV = - \int (\mathbf{M} \times \nabla) \times \mathbf{r} dV.$$

Here

$$(\mathbf{M} \times \nabla) \times \mathbf{r} = (\mathbf{M} \cdot \nabla) \mathbf{r} - \mathbf{M} (\nabla \cdot \mathbf{r}) = \mathbf{M} - 3\mathbf{M} = -2\mathbf{M}.$$

Hence, finally,

$$\int \mathbf{r} \times (\nabla \times \mathbf{M}) \, dV = 2 \int \mathbf{M} \, dV. \quad (3.15)$$

Comparing (3.15) with (3.14), we obtain

$$\int \mathbf{M} \, dV = \int \frac{\mathbf{r} \times (\rho_{\text{bound}} \mathbf{v})}{2c} \, dV. \quad (3.16)$$

Eq. (3.16) shows that the vector \mathbf{M} introduced by the relation (3.13) represents the mean magnetic moment per unit volume of the body produced by moving bound charges (see §22 of Part I).

The statistical theory of magnetic phenomena will be expounded in ch. 6.

By means of the quantities \mathbf{j} , \mathbf{P} and \mathbf{M} the mean current density in matter can finally be written in the form

$$\overline{\rho \mathbf{v}} = \mathbf{j} + \frac{\partial \mathbf{P}}{\partial t} + c(\nabla \times \mathbf{M}). \quad (3.17)$$

§4. The system of equations for the electromagnetic field in a medium

Having calculated the mean values of the quantity $\rho \mathbf{v}$, we can pass on to the final formulation of the equations for the electromagnetic field in a medium. Substituting (3.17) into (1.16), we find

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{j} + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \frac{\partial \mathbf{P}}{\partial t} + 4\pi(\nabla \times \mathbf{M}),$$

or

$$\nabla \times (\mathbf{B} - 4\pi \mathbf{M}) = \frac{4\pi}{c} \mathbf{j} + \frac{1}{c} \frac{\partial}{\partial t} (\mathbf{E} + 4\pi \mathbf{P}). \quad (4.1)$$

We now introduce a new notation, writing

$$\mathbf{H} = \mathbf{B} - 4\pi \mathbf{M} \quad (4.2)$$

and

$$\mathbf{E} = \mathbf{D} - 4\pi \mathbf{P}. \quad (4.3)$$

We rewrite (4.1) in the form

$$\nabla \times \mathbf{H} = \frac{4\pi}{c} \mathbf{j} + \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t}. \quad (4.4)$$

We now consider eq. (1.17). By expressing the mean charge density according to formula (2.6) in terms of the mean density of the free charge ρ and the mean density of the bound charge ρ_{bound} , and making use of (2.6), we obtain

$$\nabla \cdot \mathbf{E} = 4\pi\rho + 4\pi\rho_{\text{bound}} = 4\pi\rho - 4\pi(\nabla \cdot \mathbf{P}).$$

Hence, by virtue of (4.3),

$$\nabla \cdot \mathbf{D} = 4\pi\rho. \quad (4.5)$$

The two remaining equations (1.14) and (1.15), as we have seen, are averaged without any difficulties:

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad (4.6)$$

$$\nabla \cdot \mathbf{B} = 0. \quad (4.7)$$

In this case ρ and \mathbf{j} are related by the equation of continuity

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0. \quad (4.8)$$

Eqs. (4.4) – (4.8) form the system of equations for the field in a medium. This system of equations was established by Maxwell in 1873 and they are called Maxwell's equations (as distinct from the Maxwell–Lorentz equations (1.1) – (1.5) of Part I). It is clear that this system is still not complete, since it contains four unknown vectors: the mean fields \mathbf{E} and \mathbf{B} and the auxiliary quantities \mathbf{H} and \mathbf{D} . For historical reasons we call the vector \mathbf{H} the magnetic field in the medium, and the vector \mathbf{D} the electric displacement.

In order that the Maxwell system of equations may be complete, it is necessary to give certain additional relations connecting the basic and auxiliary field vectors. These relations are called constitutive equations. In the simplest (homogeneous and isotropic) medium which we have considered above, the form of the constitutive equations follows directly from (3.2).

By virtue of (3.13) and (4.2), we have

$$\mathbf{H} = \mathbf{B} - (4\pi\alpha)\mathbf{B} = [1 - (4\pi\alpha)]\mathbf{B}. \quad (4.9)$$

For historical reasons the constitutive equation (4.9) is written in the form

$$\mathbf{B} = \mu\mathbf{H}, \quad (4.10)$$

where the constant $\mu = [1 - (4\pi\alpha)]^{-1}$ is called the magnetic permeability of the medium. The induced magnetic moment per unit volume \mathbf{M} is expressed not in terms of \mathbf{B} but in terms of \mathbf{H} according to the formula

$$\mathbf{M} = (\mu\alpha)\mathbf{H} = \kappa\mathbf{H}, \quad (4.11)$$

where κ is called the magnetic susceptibility. It follows from (4.2) and (4.11) that

$$\mu = 1 + 4\pi\kappa. \quad (4.12)$$

Analogously, from (3.7) and (4.3) we obtain the constitutive equation

$$\mathbf{D} = \epsilon\mathbf{E}, \quad (4.13)$$

where the coefficient

$$\epsilon = 1 + 4\pi\chi \quad (4.14)$$

is called the dielectric permittivity or, according to an obsolete but frequently applied terminology, the dielectric constant. The relation (3.3) is also usually referred to as a constitutive equation. We would like to stress that, in contrast to the Maxwell–Lorentz equations, which are among the most accurate and universal of the known laws of nature, Maxwell's equations have limited applicability in consequence of the limited region of applicability of the constitutive equations. We shall dwell on this problem in more detail below.

In what follows it will be shown that for all bodies the electric susceptibility χ is larger than 0 and, consequently, the dielectric constant ϵ is larger than 1. On the contrary, the magnetic susceptibility can be either positive or negative.

Substances for which $\kappa > 0$ are called paramagnetic, while those for which

$\kappa < 0$ are called diamagnetic. For what follows we shall need the integral form of Maxwell's equations, which we can write as follows:

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\frac{1}{c} \frac{\partial}{\partial t} \int \mathbf{B} \cdot d\mathbf{S}, \quad (4.15)$$

$$\oint \mathbf{B} \cdot d\mathbf{S} = 0, \quad (4.16)$$

$$\oint \mathbf{H} \cdot d\mathbf{l} = \frac{4\pi}{c} \int \mathbf{j} \cdot d\mathbf{S} + \frac{1}{c} \frac{\partial}{\partial t} \int \mathbf{D} \cdot d\mathbf{S}, \quad (4.17)$$

$$\oint \mathbf{D} \cdot d\mathbf{S} = 4\pi \int \rho \, dV. \quad (4.18)$$

In just the same way as has been done for the electromagnetic field in vacuum in §10 of Part I, one can introduce the electromagnetic potentials φ and \mathbf{A} in matter. We define them by the formulae

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad (4.19)$$

$$\mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} - \nabla \varphi. \quad (4.20)$$

In this case the Lorentz conditions assume the form

$$\nabla \cdot \mathbf{A} = -\frac{\epsilon\mu}{c} \frac{\partial \varphi}{\partial t}. \quad (4.21)$$

Reproducing the calculations of §10 of Part I, one can easily arrive at the equations for the potentials

$$\nabla^2 \mathbf{A} - \frac{\epsilon\mu}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\frac{4\pi\mu}{c} \mathbf{j}, \quad (4.22)$$

$$\nabla^2 \varphi - \frac{\epsilon\mu}{c^2} \frac{\partial^2 \varphi}{\partial t^2} = -\frac{4\pi}{c} \rho. \quad (4.23)$$

§5. Boundary conditions

In what follows we shall usually have to consider electromagnetic phenomena in bodies which are limited in space. Hence it is necessary to find out how the vectors of the electromagnetic field change near the boundary of the body. In the most general case the boundary of the body is the interface of two media with different properties. We shall assume that with a sufficient degree of accuracy the interface can be considered as a surface, and we shall not be interested in the properties of the electromagnetic field in the transition layer near the boundary.

Let one medium be characterized by ϵ_1 and μ_1 , and the other by ϵ_2 and μ_2 . The behaviour of the vectors of the field at the interface can be established from Maxwell's equations written in integral form.

Let us consider eq. (4.15) and apply it to an infinitesimal contour L shown in fig. IV.1, assuming the length l_1 to be an infinitesimal quantity of the first order, and the length l_2 to be an infinitesimal quantity of the second order. Then we have

$$\oint \mathbf{E} \cdot d\mathbf{l} = E_{\text{tg}}^{(1)} l_1 - E_{\text{tg}}^{(2)} l_1 + \text{an infinitesimal quantity of the second order}.$$

Here E_{tg} denotes the component of the electric field strength vector tangent to the interface; the indices refer to the first and second media.

Eq. (4.15) gives:

$$(E_{\text{tg}}^{(1)} - E_{\text{tg}}^{(2)}) l_1 = -\frac{1}{c} \frac{\partial \Phi}{\partial t}, \quad (5.1)$$

where Φ is the flux of magnetic induction through the surface enclosed by the contour L . Obviously, $\Phi \sim l_1 l_2$ and is of a high order of small quantities. Hence, passing over to the limit $l_2 \rightarrow 0$, we find from (5.1)

$$E_{\text{tg}}^{(1)} = E_{\text{tg}}^{(2)}. \quad (5.2)$$

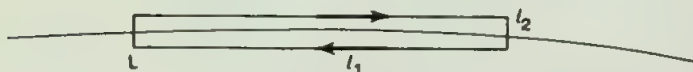


Fig. IV.1

The tangential component of the electric field strength remains continuous in passing through the interface of the media.

We apply an analogous method to formula (4.17), dropping at once the term of second order in small quantities $\oint \mathbf{D} \cdot d\mathbf{S}$. We then find

$$\oint \mathbf{H} \cdot d\mathbf{l} = (\mathbf{H}_2 - \mathbf{H}_1) \cdot \mathbf{l} dl_1 = \frac{4\pi}{c} \int \mathbf{j} \cdot d\mathbf{S},$$

where \mathbf{l} is the unit vector of the contour lying in the plane of the interface (fig. IV.1).

We denote the normal to the interface by \mathbf{n} , and the normal to the surface enclosed by the contour \mathbf{l} by \mathbf{n}_l . The vectors \mathbf{n} , \mathbf{n}_l and \mathbf{l} form a right-handed screw system

$$\mathbf{l} = \mathbf{n}_l \times \mathbf{n}.$$

Further, we introduce the concept of surface current density \mathbf{j}_S , by which is meant the amount of electricity passing per second through unit length on the surface

$$\lim_{dl_1 dl_2 \rightarrow 0} (\mathbf{j} \cdot \mathbf{n}_l) dl_1 dl_2 = (\mathbf{j}_S \cdot \mathbf{n}) dl_1.$$

Then (4.17) gives

$$\oint \mathbf{H} \cdot d\mathbf{l} = \frac{4\pi}{c} (\mathbf{j}_S \cdot \mathbf{n}_l) dl_1 = (\mathbf{H}_2 - \mathbf{H}_1) \cdot \mathbf{l} dl_1,$$

or

$$(\mathbf{H}_2 - \mathbf{H}_1) \cdot \mathbf{l} = \frac{4\pi}{c} (\mathbf{j}_S \cdot \mathbf{n}_l).$$

Expressing \mathbf{l} in terms of \mathbf{n}_l , we have

$$(\mathbf{H}_2 - \mathbf{H}_1) \cdot (\mathbf{n}_l \times \mathbf{n}) = \mathbf{n}_l \cdot [\mathbf{n} \times (\mathbf{H}_2 - \mathbf{H}_1)],$$

hence

$$\mathbf{n}_l \cdot [\mathbf{n} \times (\mathbf{H}_2 - \mathbf{H}_1)] = \frac{4\pi}{c} (\mathbf{j}_S \cdot \mathbf{n}_l).$$

Since the orientation of the vector \mathbf{n}_l in the plane of the interface can be arbitrary, the following equality must be fulfilled:

$$\mathbf{n} \times (\mathbf{H}_2 - \mathbf{H}_1) = \frac{4\pi}{c} \mathbf{j}_S. \quad (5.3)$$

In the presence of surface currents the tangential component of the field strength \mathbf{H} is discontinuous at the interface of the media. The value of the discontinuity ΔH_{tg} is equal to $(4\pi/c)j_S$.

If there is no surface current at the interface, $j_S = 0$, then

$$H_{tg}^{(1)} = H_{tg}^{(2)}. \quad (5.4)$$

The boundary conditions for the normal components of the induction vector \mathbf{B} and the displacement vector \mathbf{D} are obtained from (4.16) and (4.18), if the infinitesimal surface S shown in fig. IV.2 is chosen as the integration surface. The area S_1 of the bases is an infinitesimal quantity of the first order, while the area of the lateral faces is an infinitesimal quantity of the second order. From (4.16) we find

$$(B_n^{(1)} - B_n^{(2)}) + \text{an infinitesimal quantity of the second order} = 0,$$

or, passing over to the limit,

$$B_n^{(1)} = B_n^{(2)}. \quad (5.5)$$

The normal component of the magnetic induction vector is conserved in passing through the interface.

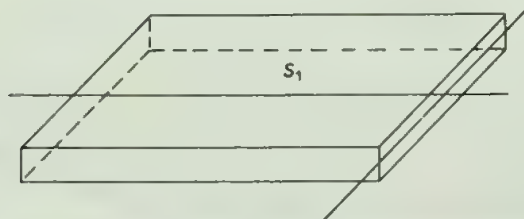


Fig. IV.2

Analogously, from (4.18) we find

$$D_n^{(1)} = D_n^{(2)} + 4\pi\omega_S, \quad (5.6)$$

where ω_S is the surface charge density defined, as above, by

$$\lim_{h \rightarrow 0} \int \rho \, dV = \int \omega_S \, dS.$$

If the surface charge density is $\omega_S = 0$, then

$$D_n^{(1)} = D_n^{(2)}. \quad (5.7)$$

Eqs. (5.2), (5.3), (5.5) and (5.6) are the boundary conditions which must be satisfied by the field vectors at the interface of the media. In particular, at the boundary with a vacuum one must assume in these formulae that $\epsilon_2 = \mu_2 = 1$.

§6. The limits of applicability of the system of constitutive equations

We have obtained Maxwell's equations (together with the constitutive equations) from the Maxwell-Lorentz equations by means of averaging, based on certain assumptions about the properties of the medium.

Although these assumptions have a rather general character, they are not always fulfilled. It turns out that the region of applicability of the constitutive equations and, consequently, Maxwell's equations in the simplest form (4.4) – (4.7) is rather limited. We have to stress these limitations particularly because nowadays one has to deal more and more often in physics with systems to which Maxwell's equations in the form we have written down are not applicable. Usually one of the terms on the right-hand side of eq. (4.4) is large while the other is small. Thus, in ideal dielectrics the conduction current is $\mathbf{j} = 0$, and in real dielectrics it is very small in comparison with the displacement current $\partial \mathbf{D} / \partial t$. On the other hand, in conductors the displacement current is usually small. Therefore eq. (4.17) should be considered as a general expression which includes limiting real cases but which is almost always simplified in actual use. However, they are presented in all the old and many new textbooks without adequate reservations.

We now go on to the discussion of the assumptions made in deriving Maxwell's equations. The basic assumption is the expansion (3.2). This ex-

pansion means, first of all, that the medium is isotropic, so that the connection between the vector $\bar{\rho}\bar{\mathbf{v}}$ and the vectors \mathbf{E} , $\partial\mathbf{E}/\partial t$ and $\mathbf{V} \times \mathbf{B}$ is characterized by the scalar constants σ , χ , μ . It is clear that this assumption is not applicable to anisotropic media, in particular to monocrystals. Hence in anisotropic media the relations (3.3), (4.13) and (4.10) do not hold and must be replaced by the tensor expressions

$$j_i = \sigma_{ik} E_k, \quad D_i = \epsilon_{ik} E_k, \quad B_i = \mu_{ik} H_k. \quad (6.1)$$

The quantities ϵ_{ik} and μ_{ik} are symmetric second-rank tensors. The proof of this statement will be given in §12.

Further, the very existence of a linear relation between \mathbf{M} and \mathbf{B} or, what is the same, between \mathbf{B} and \mathbf{H} , turns out not to be fulfilled for a relatively small but very important group of substances which are called ferromagnetics and antiferromagnetics. For these substances the functional relationship between \mathbf{B} and \mathbf{H} has a complex non-linear and ambiguous character. We shall dwell on the theory of ferromagnetism below, cf. §20.

Another group of substances, called ferroelectrics, possess electric properties which are analogous to the magnetic properties of ferromagnetics: the dependence of \mathbf{D} on \mathbf{E} is non-linear and ambiguous.

Thus, for ferromagnetics and ferroelectrics the expression (3.2), in which the first terms of the expansion are retained, loses its meaning.

A number of substances, called superconductors, show a profound change in their magnetic properties at low temperatures. The magnetic field does not penetrate inside superconductors, so that in the volume of a superconductor the following condition is fulfilled:

$$\mathbf{B} = 0.$$

It is obvious that Maxwell's equations in the form (4.4)–(4.7) do not describe the behaviour of superconductors. The properties of superconductors will be considered in §21, while the theory of superconductivity will be given in Part V.

The problem of the applicability of the expansion (3.2) in the case of high-frequency fields and for spatially non-uniform systems is more complex.

In considering high frequencies the division of the current into two parts, the current of free charges and the current of bound charges, ceases to have any meaning.

It is clear that in a high-frequency field the free charges as well as the bound charges perform practically the same oscillatory motion. Reasoning

further quantitatively, one can say that the expansion (3.2) assumes that the variation of fields in space must be rather smooth and occur over distances which are large in comparison with molecular dimensions. Otherwise the macroscopic averaging over the volume, which must include a sufficiently large number of molecules, would be meaningless. Since the variation in space takes place over a wavelength, the inequality $\lambda \gg a$ must be fulfilled and, consequently, we are confined to frequencies which are much lower than $\omega_0 \ll c/a$.

However, this restriction is still not the strongest one. In the range $\omega \ll \omega_0$ there is a wide interval of frequencies in which the constitutive equation (4.13) turns out to be inapplicable. This equation, as well as (3.2), assumes that the frequency of the field is small compared with the inverse of the relaxation time τ^{-1} which is characteristic of the given substance. In this case the polarization \mathbf{P} at a given point of space at a particular time t is determined by the displacement (mean field) at the same point and at the same instant of time. But if the frequency is comparable with the inverse of the relaxation time τ^{-1} , then the polarization will lag behind the field and will become dependent on the history of the process. Then instead of (4.13) we shall have to write the more general relation

$$\mathbf{D}(\mathbf{r}, t) = \int_{-\infty}^t \epsilon(t, t') \mathbf{E}(\mathbf{r}, t') dt', \quad (6.2)$$

where the integration is carried out with respect to past times ($t' < t$) and the function $\epsilon(t, t')$ determines the production of polarization as a result of the action of the field at preceding instants of time. Since time flows uniformly, there is no unique instant of time. This means that the function $\epsilon(t, t')$ can depend only on the time which passed between the instant t' and the instant t , in other words, on the argument $t - t'$. Thus

$$\mathbf{D}(\mathbf{r}, t) = \int_{-\infty}^t \epsilon(t - t') \mathbf{E}(\mathbf{r}, t') dt'. \quad (6.3)$$

We note that the frequencies for which eq. (4.13) is inapplicable are very small in comparison with ω_0 . Indeed, relaxation times τ are always in order of magnitude equal to characteristics atomic times $\tau_{at} \sim a/v$, where v is the velocity of electrons in atoms. Hence

$$\omega \sim \frac{1}{\tau_{at}} \sim \frac{v}{a} \sim \frac{v}{c} \omega_0 \ll \omega_0. \quad (6.4)$$

In §32 we will show that in the frequency range approaching characteristic atomic frequencies the dielectric constant is dependent on the frequency. This phenomenon is therefore called frequency dispersion or time dispersion.

Another phenomenon, called spatial dispersion, also occurs at high frequencies. In order that eq. (4.13) may apply the wavelength must be large not only in comparison with the size of the atom but also in comparison with the size of any region of spatial non-uniformity in the substance.

If the size of a region of non-uniformity is denoted by l , then it is necessary that the inequality $\lambda > l$ be satisfied. If $\lambda \approx l$, then the polarization at a given point of space will depend on the value of the field at neighbouring points of space at preceding instants of time, i.e.

$$\mathbf{D}(\mathbf{r}, t) = \int_{-\infty}^t dt' \int \epsilon(\mathbf{r}, \mathbf{r}', t-t') \mathbf{E}(\mathbf{r}', t') dV'. \quad (6.5)$$

Formula (6.5) means that a contribution to the polarization at a given point at a given instant of time is produced by charges which have earlier been at neighbouring points of space. This phenomenon is called spatial dispersion. Media with a spatial dispersion are encountered in a number of important cases (e.g.: plasma, metals). All that has been said about the time dispersion and spatial dispersion of dielectric permittivity also holds for magnetic permeability.

We thus see that the way of writing the constitutive equations and, consequently, Maxwell's equations as presented in §4 is the simplest form and is applicable only to the case of isotropic media which do not display a time dispersion or spatial dispersion and are neither ferromagnetic nor ferroelectric.

§7. The law of conservation of energy in a medium

The derivation of the law of conservation of energy for a non-absorbing, isotropic and non-ferromagnetic medium does not differ from the analogous calculations carried out in §12 of Part I. Namely, if all bodies (conductors and dielectrics) which are placed in a field are at rest, then the work done by the electric field on the charges per unit time is equal to

$$\begin{aligned} \frac{dW}{dt} &= \int \mathbf{j} \cdot \mathbf{E} dV = \frac{c}{4\pi} \int \mathbf{E} \cdot (\nabla \times \mathbf{H}) dV - \frac{1}{4\pi} \int \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} dV = \\ &= \frac{c}{4\pi} \int [\mathbf{E} \cdot (\nabla \times \mathbf{H}) - \mathbf{H} \cdot (\nabla \times \mathbf{E})] dV - \frac{1}{4\pi} \int \left(\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} \right) dV = \end{aligned}$$

$$= - \oint \boldsymbol{\sigma} \cdot d\mathbf{S} - \int \frac{\partial}{\partial t} u_0 dV.$$

Here we have added an expression equal to zero to the integral to be calculated by virtue of (1.14), and have denoted the energy density of the electromagnetic field in the medium by u_0

$$u_0 = \frac{\epsilon E^2 + \mu H^2}{8\pi} \quad (7.1)$$

and the Poynting vector by $\boldsymbol{\sigma}$

$$\boldsymbol{\sigma} = \frac{c}{4\pi} \mathbf{E} \times \mathbf{H} = \frac{c}{4\pi\mu} \mathbf{E} \times \mathbf{B}. \quad (7.2)$$

It is convenient to write the energy conservation law in the form

$$- \frac{\partial}{\partial t} \int u_0 dV = \oint \boldsymbol{\sigma} \cdot d\mathbf{S} + \frac{dW}{dt}. \quad (7.3)$$

In differential form the energy conservation law is expressed by the relation

$$- \frac{\partial}{\partial t} \frac{\epsilon E^2 + \mu H^2}{8\pi} = \mathbf{j} \cdot \mathbf{E} + \nabla \cdot \boldsymbol{\sigma}. \quad (7.4)$$

The interpretation of the conservation law for a medium does not, in principle, differ from that given in ch. 1 for a vacuum. All one should note is the fact that (7.4) contains only the free charge current density \mathbf{j} . As free charges are moving in a conductor at rest the entire mechanical work of the field is transformed completely into heat, called the Joule heat. The mechanism of this process depends on the actual properties of the conductor. We shall touch upon this problem later. Denoting the Joule heat released in unit volume per second by Q_0 equal to

$$Q_0 = \mathbf{j} \cdot \mathbf{E} = j^2/\sigma,$$

we can write the law of energy conservation in the form

$$- \frac{\partial}{\partial t} \int u_0 dV = \int Q_0 dV + \oint \boldsymbol{\sigma} \cdot d\mathbf{S}.$$

The quantities

$$\frac{\epsilon E^2}{8\pi} = \frac{\mathbf{D} \cdot \mathbf{E}}{8\pi} \quad \text{and} \quad \frac{\mu H^2}{8\pi} = \frac{\mathbf{B} \cdot \mathbf{H}}{8\pi}$$

represent respectively the densities of the electric and magnetic energies.

Since, however, the quantities ϵ and μ in matter are functions of temperature (see ch. 6), our overall treatment assumes the constancy of the temperature of the medium. Hence the quantity $(\mathbf{E} \cdot \mathbf{D} + \mathbf{B} \cdot \mathbf{H})/8\pi$ must be interpreted as the free energy per unit volume of the medium.

Electrostatics

§8. The electrostatic field

The equations of the electromagnetic field are greatly simplified in the case of electrostatics.

Assuming that the fields do not depend on time and that the current is equal to zero, one can write Maxwell's equations in the form

$$\nabla \times \mathbf{E} = 0, \quad \nabla \cdot \mathbf{D} = 4\pi\rho, \quad (8.1)$$

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{H} = 0. \quad (8.2)$$

We shall not consider constant magnetic fields in matter *.

Eqs. (8.1) and (8.2) determine the electrostatic field in a medium completely.

It should be noted that ρ denotes the mean density of free charges which are introduced into the dielectric from outside. In contrast to conductors, free charges in a dielectric are at rest and fixed in definite positions.

* See I.E. Tamm, *Osnovy teorii elektrichestva (Introduction to the theory of electricity)* (Gostekhizdat, Moscow, 1954).

If a dielectric is placed in an external electric field, the bound charges contained in the molecules, or ions forming the ion lattice, are displaced with respect to each other in such a way that the dielectric becomes polarized. For non-uniform polarization, a bound volume charge arises in the dielectric whose mean density is determined by formula (3.6). We recall that the total bound charge which appears in the polarized dielectric is always equal to zero.

In addition to the volume density of bound charges, a surface charge density $\omega_S = P_n$ arises at the surface of a polarized dielectric.

The field inside a dielectric is, according to (8.1), irrotational. The electrostatic field potential, defined by (6.2), satisfies the equation

$$\nabla \cdot (\epsilon \nabla \varphi) = -4\pi\rho. \quad (8.3)$$

In a homogeneous dielectric $\epsilon = \text{const}$, and the last equation becomes the Poisson equation

$$\nabla^2 \varphi = -4\pi\rho/\epsilon. \quad (8.4)$$

From this equation it follows that for a homogeneous dielectric the potential of the field produced by the free charge volume density ρ can be written in the form

$$\varphi = \frac{1}{\epsilon} \int \frac{\rho \, dV}{r}. \quad (8.5)$$

Formula (8.5) shows that the potential of the field and the field itself in a dielectric medium are decreased by a factor of ϵ in comparison with the field produced by the same charges in vacuum. In particular, the field of a point charge in a dielectric medium is equal to

$$\mathbf{E} = \frac{e\mathbf{R}}{\epsilon R^3}.$$

The corresponding force of interaction of charges according to the Coulomb law is smaller in a homogeneous dielectric medium than in vacuum by a factor ϵ . This result has a simple meaning: the electric field produced by free charges which are put in a dielectric medium gives rise to its polarization. As a result of the polarization the bound charges are displaced and produce in the medium a field which decreases the field of the free charges.

It should be stressed that this refers only to the field of charges which are put inside a homogeneous and isotropic dielectric. From examples which will be discussed below it will be clear that this conclusion is quite inapplicable to an inhomogeneous dielectric.

At the interface of two dielectrics the potential must satisfy the following boundary conditions:

$$\varphi_1 = \varphi_2, \quad (8.6)$$

$$\epsilon_1 \left(\frac{\partial \varphi}{\partial n} \right)_1 = \epsilon_2 \left(\frac{\partial \varphi}{\partial n} \right)_2. \quad (8.7)$$

The first of these is equivalent to (5.2), while the second follows directly from (5.6) if the surface density of free charges is assumed to be $\omega_s = 0$.

Representing the field distribution in an obvious way by means of lines of force, we see from the boundary conditions (8.6) and (8.7) that the lines of force are broken at the interface of two dielectrics and that

$$\frac{\operatorname{tg} \alpha_1}{\operatorname{tg} \alpha_2} = \frac{\epsilon_1}{\epsilon_2}. \quad (8.8)$$

The meaning of the angles α_1 and α_2 is shown in fig. IV.3.

Let us now consider the electrostatic field in conductors. As we have already stressed, a distinctive feature of conductors is the presence of a large number of mobile charges in them. The presence of a mean field \mathbf{E} inside a conductor always produces an electric current in it, representing the motion of free charges under the action of the field. If current is absent in a conduc-

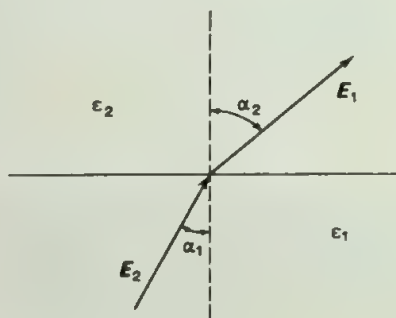


Fig. IV.3

tor, and this is just the requirement of electrostatics, then inevitably the field strength in the conductor is also equal to zero.

$$\mathbf{E} = 0.$$

From the equality of the field to zero in the entire region of space occupied by the conductor it follows that the volume charge density inside the conductor is also equal to zero.

The reduction to zero of the electrostatic field strength vector in the conductor can be presented in the following obvious way. When a conductor is placed in an external field, free charges in it begin to move towards the surface, and this displacement continues until the field of the surface charges completely compensates for the field inside the conductor due to external sources.

The process proceeds as though the conductor possesses an infinitely large polarizability for any value of the field strength or, as is seen from (8.7), an infinitely large dielectric constant $\epsilon_2 \rightarrow \infty$. As is seen from (8.8), the lines of the electrostatic field are normal to the surface of the conductor. Although the identification of the conductor with a dielectric having an infinitely large permittivity has a formal character, the concept of polarization is not applicable to free charges in conductors and is only useful for an obvious interpretation of the relations of electrostatics.

The value of the surface charge density for conductors is obtained from the boundary condition (5.6), if the field strength inside the metal is assumed to be equal to zero. Then we find

$$\omega_S = \epsilon E_n / 4\pi, \quad (8.9)$$

where ϵ is the dielectric constant of the medium surrounding the conductor, and E_n is the normal component of the external field at its surface. The boundary condition (5.2) gives for the tangential component of the external field at the surface of the conductor

$$E_{tg} = 0, \quad (8.10)$$

since in the conductor itself it is equal to zero.

Thus, the field external to the conductor is directed perpendicularly to its surface and $E_n = E$. Formulae (8.9) and (8.10) can be rewritten by introducing the external field potential φ in the form

$$\left. \begin{aligned} \omega_S &= -\frac{\epsilon}{4\pi} \frac{\partial \varphi}{\partial n}, \\ \varphi &= \varphi_m = \text{const} \end{aligned} \right\} \text{ at the surface of the conductor.} \quad (8.11)$$

In particular, if the metal is in vacuum, the value of ϵ must be set equal to unity.

The total charge at the surface of the conductor is

$$e = -\frac{\epsilon}{4\pi} \oint_S \left(\frac{\partial \varphi}{\partial n} \right) dS, \quad (8.12)$$

where the integration is carried out over the entire surface S of the conductor. Formula (8.12) establishes the relation between the surface potential φ_m of the conductor and its charge e . It is easy to verify that this relation has a linear character: if φ is increased by a factor k , then by virtue of (8.12) and the constancy of the potential along the surface of the metal its charge e will increase by the same factor.

The ratio of the charge of a conductor to its potential is called its capacitance:

$$C = \frac{e}{\varphi_m} = \frac{\epsilon}{4\pi} \oint_S \frac{1}{\varphi_m} \left(\frac{\partial \varphi}{\partial n} \right) dS. \quad (8.13)$$

The capacitance is proportional to the dielectric constant ϵ of the medium, and in other respects is determined solely by the form of the conductor.

As is seen from (8.13), the capacitance has the dimensions of length, and in order of magnitude is the same as the linear dimension of the conductor. In the case where there is not one but several conductors in the electrostatic field, the boundary conditions (8.13) must be fulfilled for each of these.

In this case the charges on the conductors and their potentials are not arbitrary. There is a certain relation between them which follows from the uniqueness of the solution of the Laplace equation with boundary conditions of the type (8.11). From the mathematical standpoint finding the influence of one conductor on the others is a complex problem, and we shall not touch upon it *.

* See, for example, W.R.Smythe, *Static and dynamic electricity* (McGraw-Hill, New York, 1950).

§9. The solution of electrostatic problems

We can now formulate the problem of electrostatics in a general way.

Let a system of conductors and dielectrics of different natures be given in a finite region of space. Let the volume charge density ρ at all points of space also be given (in the volume occupied by the conductors $\rho = 0$). Then the equation for the potential has the form

$$\nabla^2 \varphi = -4\pi\rho/\epsilon_i$$

(i corresponds to the medium; in vacuum $\epsilon = 1$).

At the interfaces vacuum–dielectric, vacuum–conductor and dielectric–conductor of the media the system of boundary conditions which have been discussed is fulfilled. The potential at infinity satisfies the condition

$$\varphi \sim O\left(\frac{1}{r^{1+\alpha}}\right) \quad (\alpha > 0). \quad (9.1)$$

We wish to find the potential distribution and electric field distribution over all space. This problem is often called the direct problem of electrostatics. The inverse problem is to find the charge distribution from the known potential distribution. The direct problem of electrostatics is one of the basic problems of mathematical physics. The electrostatic problem can be solved by elementary methods only for particular simple cases. A number of methods for its solution have been devised, and solutions for a great number of actual systems have been calculated. In these an important role is played by approximate and numerical methods. Our purpose does not include the analysis of different cases of electrostatic fields. We shall consider only some simple examples of general interest.

The solution of Poisson's or Laplace's equations for given boundary conditions represents one of the most important problems of mathematical physics.

Standard methods, exact as well as approximate, have been devised for the solution of these problems. The solution of the boundary value problems of electrostatics is of great practical importance, but is of no great interest from the point of view of theoretical physics. Hence, we refer the reader to handbooks of mathematical physics and we shall confine ourselves only to the briefest exposition.

First of all, we point out the existence of two types of boundary conditions. If the region in which the potential is to be found is bounded by

metal conductors at the surface of which the potential (i.e. the function sought for) is defined, then it is said that Dirichlet's boundary conditions are defined. But if the boundary surface is a conductor with a given surface charge density, then according to (8.11) the derivative of the potential with respect to the normal is defined. In this case the boundary conditions are called Neumann's conditions.

The definition of the Laplace (or Poisson) equation together with the boundary conditions at all boundary surfaces and at infinity defines the boundary value problem.

It can easily be shown that the boundary value problem with Dirichlet's or Neumann's conditions has a unique solution.

We shall dwell very briefly on the solution of boundary value problems for Laplace's equation, when volume charges are absent.

There are a number of methods for the solution of such equations. We shall mention only some of these.

If the boundary value problem has a particular simple symmetry, then the Laplace operator is most conveniently written in coordinates which express the symmetry of the system. Thus, for example, to find fields with a spherical or cylindrical symmetry one has to express the Laplace operator in spherical or cylindrical coordinates.

In such cases the variables are separable and the solution of the equation with partial derivatives reduces to the solution of equations in total derivatives. There are eleven different coordinate systems in which the Laplace operator in three dimensions allows one to separate the variables. They are given in the book of P.M.Morse and H.Feshbach *.

As the simplest examples we shall consider the solutions of Laplace's equations in Cartesian coordinates. In these coordinates Laplace's equation has the form:

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2} = 0. \quad (9.2)$$

Assume that the boundary conditions are given at surfaces which have in Cartesian coordinates a simple form, for example,

* P.M.Morse and H.Feshbach, *Methods of theoretical physics* (McGraw-Hill, New York, 1953).

$$\begin{array}{lll}
 \varphi = \varphi_1(x, y) & \text{for} & z = 0, \quad x, y \leq L, \\
 \varphi \rightarrow 0 & & z \rightarrow \infty, \\
 \varphi = 0 & & x = \pm L, \quad y = \pm L.
 \end{array} \quad (9.3)$$

These conditions mean that at the surface of an open parallelepiped the potential is equal to zero at the lateral walls and is defined at the base.

Then eq. (9.2) is easily solved by the method of separation of variables. We shall seek the solution of (9.2) in the form of the product

$$\varphi = X(x)Y(y)Z(z), \quad (9.4)$$

where X , Y , Z are functions of one variable. The substitution of (9.4) leads to the equation

$$X \frac{d^2 X}{dx^2} + \frac{1}{Y} \frac{d^2 Y}{dy^2} + \frac{1}{Z} \frac{d^2 Z}{dz^2} = 0. \quad (9.5)$$

Since each of the terms depends on its argument the equality (9.5) is fulfilled if they all are equal to constants, i.e.

$$\frac{1}{X} \frac{d^2 X}{dx^2} = -\alpha^2, \quad \frac{1}{Y} \frac{d^2 Y}{dy^2} = -\beta^2, \quad \frac{1}{Z} \frac{d^2 Z}{dz^2} = \gamma^2,$$

$$\alpha^2 + \beta^2 = \gamma^2.$$

Consequently, the solution of (9.4) has the form

$$\varphi = A (\sin \alpha x \sin \beta y) e^{-\gamma z}.$$

The boundary conditions (9.3) are fulfilled if

$$\alpha = \frac{n\pi}{L}, \quad \beta = \frac{m\pi}{L}, \quad \gamma = \frac{\pi}{L} (n^2 + m^2)^{\frac{1}{2}}; \quad n, m = 1, 2, 3, \dots,$$

so that

$$\varphi = \sum A_{nm} \sin\left(\frac{n\pi}{L} x\right) \sin\left(\frac{m\pi}{L} y\right) \exp\left[-\frac{\pi}{L} (n^2 + m^2)^{\frac{1}{2}} z\right]. \quad (9.6)$$

The coefficients A_{nm} must now be chosen in such a way that the following condition be satisfied:

$$\varphi_1(x, y) = \sum A_{nm} \sin\left(\frac{n\pi}{L} x\right) \sin\left(\frac{m\pi}{L} y\right). \quad (9.7)$$

This can always be done, since the sines form a system of orthonormal functions and by multiplying (9.7) through by $\sin(n\pi x/L) \sin(m\pi y/L)$ and integrating with respect to x and y one can write

$$A_{nm} = \frac{1}{L^2} \int_{-L}^{+L} dx \int_{-L}^{+L} \varphi_1(x, y) \sin\left(\frac{n\pi}{L} x\right) \sin\left(\frac{m\pi}{L} y\right) dy.$$

§ 10. The method of images and method of inversion

In solving a number of electrostatic problems it is useful to employ the uniqueness theorem. Indeed, suppose that for a given system we managed to choose potentials or electric fields which satisfy the differential equations and boundary conditions. By virtue of the uniqueness theorem the solution found can be considered as correct irrespective of the way in which it is obtained. As an example we consider the following problem.

Let an infinite plane $x = 0$ be the surface of a metal which occupies a half-space $x < 0$. In vacuum, at a distance a from the metal-vacuum boundary, let there be a charge e . We require to find the field strength over all space. It is obvious that the electric field inside the metal (region $x < 0$) is equal to zero.

Poisson's equation for the potential in the region $x > 0$ has the form

$$\nabla^2 \varphi = -4\pi e \delta(\mathbf{r} - \mathbf{a}). \quad (10.1)$$

The boundary conditions for our problem have the following form:

$$\varphi = 0 \quad \text{as} \quad r \rightarrow \infty.$$

In the plane $x = 0$ the following relations must be satisfied:

$$\left. \begin{aligned} D_n^{(2)} &= D_n^{(1)} + 4\pi\omega_S = 4\pi\omega_S, \\ E_{tg}^{(1)} &= E_{tg}^{(2)}. \end{aligned} \right\} \quad (10.2)$$

Here the fact that $D_n^{(1)} = 0$ in the metal and at its surface is taken into account. The condition for E_{tg} is equivalent to the requirement of the constancy of the potential at the surface of the metal. If we find the electric field outside the metal, then the boundary condition (10.2) will allow us to determine the surface charge.

To obtain the solution we note that eq. (10.1) and the condition of the constancy of the potential at the surface $x = 0$ are satisfied by the field potential of two charges, one of which (e) is located at the point $x = +a$ and the other ($-e$) at the point $x = -a$:

$$\varphi = \frac{e}{r} - \frac{e}{r_1}, \quad (10.3)$$

where r and r_1 are the distances from the charges to the point of observation. By virtue of the uniqueness of the solutions of Maxwell's equations, formula (10.3) gives the distribution to be determined.

The calculation of the gradient of φ allows one to determine the electric field E .

We now pass on to the treatment of the more complex case where the plane $x = 0$ is the interface of two dielectrics. The half-space $x > 0$ and $x < 0$ are occupied by dielectrics with dielectric constants ϵ_1 and ϵ_2 respectively. At the point $x = a$ there is a charge e . We have to find the distribution of the field over all space produced by the charge located at the point $x = a$.

The differential equation for the potential in the region $x > 0$ has the form

$$\nabla^2 \varphi = -\frac{4\pi e}{\epsilon_1} \delta(\mathbf{r}-\mathbf{a}).$$

We assume that there are no free charges at the interface, so that the boundary conditions can be written in the form

$$\varphi = 0 \quad \text{as} \quad r \rightarrow \infty,$$

$$D_n^{(2)} = D_n^{(1)}, \quad (10.4)$$

$$E_{tg}^{(2)} = E_{tg}^{(1)}. \quad (10.5)$$

We shall try to solve this problem by analogy with the preceding one, assuming that the field potential is the same as the potential of the field of equivalent charges at the points $x = a$ and $x = -a$.

We assume that the field in the region $x > 0$ is equivalent to the field of the two charges: charge e located at the point $x = a$ and an unknown charge e_1 located at the point $x = -a$. In this case the field potential in the region $x > 0$ has the form (fig. IV.4)

$$\varphi_1 = \frac{e_1}{r_1} + \frac{e}{\epsilon_1 r} \quad (x > 0). \quad (10.6)$$

Further we assume that the potential in the region $x < 0$ is also represented in the form of the field potential of a point charge

$$\varphi_2 = e_2/r, \quad (10.7)$$

where e_2 is a certain unknown charge located at the point $x = a$.

It is obvious that the potentials φ_1 and φ_2 are the solutions of the equations of electrostatics for the corresponding point charges. The electric fields in the two regions have the form

$$E_1 = \frac{e_1 \mathbf{r}_1}{r_1^3} + \frac{e \mathbf{r}}{\epsilon_1 r^3},$$

$$E_2 = \frac{e_2 \mathbf{r}}{r^3}.$$

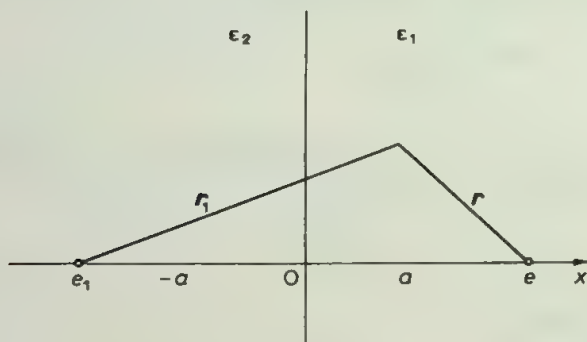


Fig. IV.4

Substituting these values into the boundary conditions (10.4) and (10.5), we find

$$\epsilon_2 \frac{e_2(\mathbf{r})_n}{r^3} = \epsilon_1 \left[\frac{e(\mathbf{r})_n}{\epsilon_1 r^3} + \frac{e_1(\mathbf{r}_1)_n}{r_1^3} \right],$$

$$\frac{e_1(\mathbf{r}_1)_{tg}}{r_1^3} + \frac{e(\mathbf{r})_{tg}}{\epsilon_1 r^3} = \frac{e_2(\mathbf{r})_{tg}}{r^3}.$$

Here $(\mathbf{r})_n$, $(\mathbf{r})_{tg}$ and $(\mathbf{r}_1)_n$, $(\mathbf{r}_1)_{tg}$ are respectively the normal and tangential projections of the vectors \mathbf{r} and \mathbf{r}_1 .

In the plane $x = 0$ the following obvious relations hold:

$$|\mathbf{r}| = |\mathbf{r}_1|, \quad (\mathbf{r}_1)_{tg} = (\mathbf{r})_{tg}, \quad (\mathbf{r}_1)_n = -(\mathbf{r})_n.$$

Hence we obtain two equations for the determination of the charges e_1 and e_2 :

$$e_2 = e_1 + (e/\epsilon_1),$$

$$e_1 \epsilon_1 = e - \epsilon_2 e_1,$$

from which it follows that

$$e_2 = \frac{2e}{\epsilon_1 + \epsilon_2},$$

$$e_1 = \frac{(\epsilon_1 - \epsilon_2)e}{\epsilon_1(\epsilon_1 + \epsilon_2)}.$$

Thus, the field potential is expressed by the formulae

$$\varphi = \frac{e}{\epsilon_1 r} + \frac{(\epsilon_1 - \epsilon_2)}{\epsilon_1(\epsilon_1 + \epsilon_2)} \frac{e}{r_1} \quad (x > 0),$$

$$\varphi = \frac{2}{\epsilon_1 + \epsilon_2} \frac{e}{r} \quad (x < 0).$$

The method of images is applied in an analogous way to obtain the solution of problems which are more complex from the geometrical point of view.

Another useful method of solving electrostatic problems is the method of inversion.

If $\varphi(r, \theta, \psi)$ is a potential satisfying the Laplace equation, then a simple calculation shows that the function

$$\varphi_1 = -\frac{R}{r} \varphi\left(\frac{R^2}{r}, \theta, \psi\right),$$

also satisfies this equation. The relation

$$\varphi_1 + \varphi = 0$$

is satisfied on a sphere of radius $r = R$. From the geometrical point of view the transformation $r \rightarrow R^2/r$ represents the mirror reflection in a sphere of radius R .

As a simple example of the application of the method of inversion we give the problem of the potential distribution in a system of a point charge and a grounded sphere of radius R . We assume the potential of the sphere to be equal to zero. If the charge e_1 is located at a distance p_1 from the centre of the sphere, then the potential of the field produced by this charge and by the charge induced on the sphere must be zero at the surface $r = R$. This condition is satisfied by the potential outside the sphere, given by the formula

$$\varphi = \frac{e_1}{r_1} - \frac{e_2}{r_2},$$

if the unknown quantities, the charge e_2 and its position, are chosen properly.

We assume that a fictitious charge is placed at a point a distance p_2 from the centre of the sphere and that

$$p_1 p_2 = R^2.$$

If the absolute value of this charge is equal to

$$e_2 = e_1 (p_1/p_2)^{\frac{1}{2}},$$

then the two charges produce a potential equal to zero at the surface of the sphere. Thus,

$$\varphi = \frac{e_1}{r_1} - \frac{R e_1}{p_2 r_2}.$$

§11. The energy of a system of conductors

We now pass on to the consideration of energy relations in the electrostatic field. The energy relations have their simplest form for conductors.

Since the field does not penetrate inside conductors, their thermodynamic properties do not change. However, the field in the space surrounding a conductor depends on the presence of the conductor, its size and so on.

We shall call the energy of the electromagnetic field surrounding a conductor the electromagnetic energy of the conductor. The total energy of a conductor is equal to the sum of its internal and electromagnetic energies.

We shall only be interested below in the latter, and for brevity we shall speak simply of the energy of the conductor.

We assume first that the conductors are in vacuum. The total energy of the electrostatic field is

$$U = \frac{1}{8\pi} \int E^2 dV.$$

We transform the integral, writing it in the form

$$\begin{aligned} U &= -\frac{1}{8\pi} \int \mathbf{E} \cdot (\nabla \varphi) dV = -\frac{1}{8\pi} \int \nabla \cdot (\varphi \mathbf{E}) dV + \frac{1}{8\pi} \int \varphi (\nabla \cdot \mathbf{E}) dV = \\ &= -\frac{1}{8\pi} \int_{R \rightarrow \infty} \varphi \mathbf{E} \cdot d\mathbf{S} + \frac{1}{8\pi} \sum_i \oint \varphi_i \mathbf{E}_i \cdot d\mathbf{S}_i. \end{aligned}$$

The integration is carried out with respect to all surfaces bounding the integration volume. These are the external surface of radius $R \rightarrow \infty$ and the surfaces of the conductors. Since the external normals of these surfaces are oriented in opposite directions, different signs occur in front of the surface integrals. We have also made use of the fact that in vacuum

$$\nabla \cdot \mathbf{E} = 0.$$

The integral with respect to the infinitely distant surface is equal to zero by virtue of (9.1). At metal surfaces $\varphi_{m_i} = \text{const}$, hence

$$\begin{aligned}
 U &= \frac{1}{8\pi} \sum_i \oint \varphi_i \mathbf{E}_i \cdot d\mathbf{S}_i = \frac{1}{8\pi} \sum_i \varphi_i \oint \mathbf{E}_i \cdot d\mathbf{S}_i = \\
 &= -\frac{1}{8\pi} \sum_i \varphi_{m_i} \oint \left(\frac{\partial \varphi_{m_i}}{\partial n} \right) dS_i = \frac{1}{2} \sum_i e_i \varphi_{m_i}.
 \end{aligned}$$

The energy of a system of conductors is formally the same as the energy of a system of charges.

We now assume that the space between the conductors is filled with a dielectric. Then two cases are, in principle, possible:

(a) the conductors are insulated, so that the charges of all the conductors have a constant value,

(b) the conductors are connected to devices which maintain their potentials constant.

In the first case, as is seen from formula (8.12), we have

$$e = -\frac{1}{4\pi} \oint \frac{\partial \varphi}{\partial n} dS = -\frac{\epsilon}{4\pi} \oint \frac{\partial \varphi'}{\partial n} dS,$$

so that

$$\varphi' = \varphi/\epsilon.$$

Correspondingly

$$U' = \frac{1}{2} \sum_i e_i \varphi_{m_i} = U/\epsilon.$$

The fields and energy are decreased by a factor of ϵ . The energy is used in the work done in filling the space with the dielectric.

In the second case, after filling the interspace with the dielectric the charge of each of the conductors will be equal to

$$e' = \frac{\epsilon}{4\pi} \oint \frac{\partial \varphi}{\partial n} dS = \epsilon e,$$

and the energy to

$$U' = \epsilon U.$$

The increase in the energy and the work done on the dielectric arise from the devices which maintain the potentials of the conductors constant.

§ 12. Dielectrics and conductors in an external electrostatic field

If an uncharged conductor or dielectric is placed in an external field $E_0^{(e)}$, the configuration of the field near the body will change. This change depends essentially on the form of the body and the character of the external field, as is clearly seen from the examples which are discussed below.

We shall consider first the case of a dielectric.

In what follows we shall assume the external field to be uniform over the extent of the body. We indicate by the index i all quantities which refer to the region of space inside the body.

As the simplest examples we shall consider a long cylinder whose axis is oriented along the field and a thin parallel-sided plate situated perpendicularly to the field.

In the first case the polarization of the dielectric will give rise to surface charges at the bases of the cylinder. However, since the length of the cylinder is large, these charges will produce only a weak field which does not distort the field due to external sources inside the cylinder. This is seen from the boundary condition (5.2) which gives

$$E_{tg}^{(i)} = E_{tg}^{(e)}$$

and, since E is, by symmetry, parallel to the axis of the cylinder,

$$E^{(i)} = E_0^{(e)}. \quad (12.1)$$

In the second example, surface charges produce an appreciable weakening of the field inside the plate. The boundary condition (5.7) gives

$$\begin{array}{l} D^{(i)} = E_0^{(e)}, \\ \text{or} \quad E^{(i)} = E_0^{(e)} - 4\pi P. \end{array} \quad (12.2)$$

The field inside the plate is reduced by an amount $4\pi P$. This inference, as is shown by calculation, is preserved qualitatively for bodies of a more complicated form.

As the next example we consider a dielectric sphere placed in a uniform

external field $E_0^{(e)}$. The potential distribution will be determined from the solution of Laplace's equation

$$\nabla^2 \varphi = 0,$$

which satisfies the requirement

$$\varphi = -E_0^{(e)} \cdot \mathbf{r} \quad (12.3)$$

at a large distance from the sphere ($r \gg R$) and satisfies the boundary conditions (8.6) – (8.7) at its surface. We seek the solution in the form

$$\varphi^{(e)} = \varphi_0 + \varphi_1^{(e)}(r, \theta, E_0^{(e)}),$$

where $\varphi_0 = -E_0^{(e)} \cdot \mathbf{r}$, and $\varphi_1^{(e)}$ represents the change in the potential near the sphere.

The function $\varphi_1^{(e)}$ must decrease for $r \gg R$. Since $\varphi_1^{(e)}$ depends only on the vector $E_0^{(e)}$ and on the scalars r and θ and must itself be a scalar, one can write it in the form of a combination of the vectors $E_0^{(e)}$ and \mathbf{r} or $E_0^{(e)}$ and the gradient of r^{-1} . In the first case $\varphi_1^{(e)}$ would not satisfy the requirement of decrease at infinity, so that the only possible form of $\varphi_1^{(e)}$ is the following:

$$\varphi_1^{(e)} = \alpha E_0^{(e)} \cdot (\nabla r^{-1}),$$

where α is a constant. Hence

$$\varphi^{(e)} = -E_0^{(e)} \cdot \mathbf{r} + \alpha E_0^{(e)} \cdot (\nabla r^{-1}). \quad (12.4)$$

On the contrary, inside the sphere the potential $\varphi^{(i)}$ at all points must remain finite and hence must be written in the form

$$\varphi^{(i)} = \beta(E_0^{(e)} \cdot \mathbf{r}). \quad (12.5)$$

It is obvious that both (12.4) and (12.5) satisfy Laplace's equation, and that φ_0 behaves at infinity in accordance with (12.3).

To find the two unknown quantities α and β one can make use of two boundary conditions:

$$\left. \begin{aligned} \varphi^{(i)} &= \varphi^{(e)} , \\ \epsilon^{(i)} \frac{\partial \varphi^{(i)}}{\partial r} &= \epsilon^{(e)} \frac{\partial \varphi^{(e)}}{\partial r} \end{aligned} \right| \quad \text{for } r = R .$$

Elementary calculations, which are most conveniently carried out in spherical coordinates, give

$$\alpha = \frac{\epsilon^{(e)} - \epsilon^{(i)}}{\epsilon^{(i)} + 2\epsilon^{(e)}} R^3 ,$$

$$\beta = - \left(1 - \frac{\epsilon^{(i)} - \epsilon^{(e)}}{\epsilon^{(i)} + 2\epsilon^{(e)}} \right) = - \frac{3\epsilon^{(i)}}{\epsilon^{(i)} + 2\epsilon^{(e)}} ,$$

so that finally

$$\varphi^{(e)} = - (\mathbf{E}_0^{(e)} \cdot \mathbf{r}) + \frac{(\mathbf{P} \cdot \mathbf{r})}{r^3} , \quad (12.6)$$

$$\varphi^{(i)} = - (\mathbf{E}_0^{(e)} \cdot \mathbf{r}) + \frac{4\pi}{3} (\mathbf{P} \cdot \mathbf{r}) , \quad (12.7)$$

where

$$\mathbf{P} = \frac{3}{4\pi} \frac{\epsilon^{(i)} - \epsilon^{(e)}}{\epsilon^{(i)} + 2\epsilon^{(e)}} \mathbf{E}_0^{(e)} , \quad (12.8)$$

$$V = \frac{4\pi}{3} R^3 .$$

In particular, in vacuum $\epsilon^{(e)} = 1$ and

$$\mathbf{P} = \frac{3}{4\pi} \frac{\epsilon^{(i)} - 1}{\epsilon^{(i)} + 2} \mathbf{E}_0^{(e)} = \frac{3\kappa}{\epsilon^{(i)} + 2} \mathbf{E}_0^{(e)} . \quad (12.9)$$

If the susceptibility κ of the body is small, so that $\epsilon^{(i)} \approx 1$, then

$$\mathbf{P} \approx \kappa \mathbf{E}_0^{(e)} .$$

Hence the field strength outside the sphere is

$$\mathbf{E}^{(e)} = -\nabla\varphi^{(e)} = \mathbf{E}_0^{(e)} + V \frac{3\mathbf{r}(\mathbf{P}\cdot\mathbf{r}) - r^2\mathbf{P}}{r^5} \quad (r < R); \quad (12.10)$$

$$\mathbf{E}^{(i)} = -\nabla\varphi^{(i)} = -\frac{4}{3}\pi\mathbf{P} + \mathbf{E}_0^{(e)} \quad (r < R). \quad (12.11)$$

In calculating the gradient according to (1.47) we have taken into account the constancy of the vector \mathbf{P} .

From formulae (12.6) – (12.11) it follows that

(1) in a uniform external field a sphere becomes polarized, acquires a dipole moment $(\mathbf{P}\cdot\mathbf{V})$, and produces an additional field identical with the field of a dipole placed at the centre of the sphere.

(2) inside the sphere the field has a constant value and is weakened by a factor of $3\epsilon^{(i)}/(\epsilon^{(i)}+2\epsilon^{(e)})$ in comparison with the external field.

Let us now consider an uncharged metal sphere placed in a uniform external field E_0 . The field outside the sphere will be found from the solution of Laplace's equation satisfying the condition (12.3) at infinity and the condition

$$\varphi = \varphi_m = 0 \quad (12.12)$$

at the surface of the metal (the potential φ_m can be assumed to be zero).

One can, however, simply assume that in the formulae for $\varphi^{(e)}$, $\mathbf{E}^{(e)}$ and \mathbf{P} , which have been obtained above for a dielectric sphere, the dielectric permittivity inside the sphere $\epsilon^{(i)}$ tends to infinity. One then obtains immediately the expression

$$\varphi^{(e)} = -E_0^{(e)}r \left(1 - \frac{R^3}{r^3}\right) \cos\theta = -\mathbf{E}_0^{(e)}\cdot\mathbf{r} + \frac{(\mathbf{r}\cdot\mathbf{P})}{r^3} V,$$

where it is assumed that

$$\mathbf{P} = \frac{3}{4\pi} \mathbf{E}_0^{(e)}.$$

The field inside the sphere is

$$\mathbf{E}^{(i)} = 0.$$

The surface charge density at the surface of the sphere is equal to

$$\omega_S = -\frac{1}{4\pi} \left(\frac{\partial \varphi}{\partial r} \right)_{r=R}^{(e)} = \frac{3E_0}{4\pi} \cos \theta .$$

The total charge of the uncharged sphere obviously remains equal to zero.

§ 13. The thermodynamic potentials of a dielectric and the dielectric susceptibility

We have seen in the preceding section that the field inside a dielectric is essentially different from the external field, and that the value of the field strength inside the dielectric depends on its shape.

Let a dielectric be placed in an external electric field of strength E_0 . For simplicity we shall restrict ourselves to the case of a uniform and isotropic dielectric. From the statistical point of view a dielectric can be considered as a quasi-closed subsystem placed in an external field of force.

The energy of such a subsystem will differ from the energy of the same subsystem without a field. The field strength is an external parameter with which the energy levels of the system change.

According to the results of § 22 of Part III, the energy of the i th level changes with the change of the parameter according to formula (22.1). In our case the external parameter is the vector E_0 . Hence its change by an amount δE_0 leads to a change in the energy of the i th level by an amount

$$\delta \epsilon_i = - \mathcal{P}_i \delta E_0 , \quad (13.1)$$

where \mathcal{P}_i represents the generalized force corresponding to the external parameter E^* :

$$\mathcal{P}_i = - \partial \epsilon_i / \partial E_0 .$$

Since we shall be interested in mean values, then, averaging (13.1), we have

$$\delta U = \overline{\delta \epsilon_i} = \overline{\mathcal{P}_i} \cdot \delta E_0 = - \overline{\mathcal{P}} \cdot \delta E_0 , \quad (13.2)$$

* The derivative with respect to the vector has the usual meaning of the abbreviated notation of the relations for vector components.

where

$$\mathcal{P} = \overline{\mathcal{P}}_i = -\partial U / \partial E_0 \quad (13.3)$$

and U is the mean energy of the system in the external field (for convenience we do not write the averaging bar and simply drop the index i).

Comparing (13.2) with (17.8) of Part I, it is easily seen that \mathcal{P} represents the mean dipole moment of the entire dielectric.

From the definition of the mean, we find

$$\begin{aligned} \mathcal{P} &= \frac{\sum P_i e^{-\epsilon_i/kT} \Omega(\epsilon_i)}{\sum e^{-\epsilon_i/kT} \Omega(\epsilon_i)} = - \frac{\sum (\partial \epsilon_i / \partial E_0) e^{-\epsilon_i/kT} \Omega(\epsilon_i)}{\sum e^{-\epsilon_i/kT} \Omega(\epsilon_i)} = \\ &= kT \frac{\partial}{\partial E_0} \ln \sum e^{-\epsilon_i/kT} \Omega(\epsilon_i) = kT \frac{\partial \ln Z}{\partial E_0} = - \frac{\partial F}{\partial E_0}, \end{aligned} \quad (13.4)$$

where F is the free energy.

The derivatives with respect to the projections of the external parameter E_0 are taken for constant values of the temperature and all other external parameters characterizing the state of the system. Formula (13.4) shows that the mean dipole moment \mathcal{P} of the body and its polarization $\mathbf{P} = \mathcal{P}/V$ can be calculated if the partition function of the system and its dependence on the strength of the external field are known. On the basis of (13.4) we can establish the relation between the change in the free energy with the field, dF , and without the field, dF_0 , in the form

$$dF = dF_0 - \mathcal{P} \cdot dE_0. \quad (13.5)$$

Integrating (13.5), we find

$$F = F_0 - \int \mathcal{P} \cdot dE_0 = F_0 - V \int \mathbf{P} \cdot d\mathbf{E}_0 = F_0 - \frac{1}{2} (\mathbf{P} \cdot \mathbf{E}_0) V. \quad (13.6)$$

We have assumed that in a uniform field and uniform dielectric medium the polarization vector $\mathbf{P} \sim \mathbf{E}_0$. This always holds for uniformly polarized bodies (in particular, bodies of ellipsoidal form).

The polarization \mathbf{P} of the body is connected with the field \mathbf{E} by the relation (3.7). Generally speaking, the field \mathbf{E} differs from the external field \mathbf{E}_0 ,

since the dielectric distorts the external field in the space surrounding it. If, however, its dielectric constant ϵ is close to unity, then $\mathbf{E}_0 \approx \mathbf{E}$ and $\mathbf{P} = \kappa \mathbf{E} \approx \kappa \mathbf{E}_0$, and

$$dF = dF_0 - V\mathbf{P} \cdot d\mathbf{E}, \quad (13.7)$$

or

$$F = F_0 - \frac{1}{2}\kappa V E^2 = F_0 - \frac{1}{2}(\mathbf{P} \cdot \mathbf{E}) V. \quad (13.8)$$

As well as the free energy F of a dielectric in an external field one often considers the total free energy of the dielectric allowing for its potential energy in the external field

$$dF' = dF_0 + V\mathbf{E}_0 \cdot d\mathbf{P} = dF_0 + V\mathbf{E} \cdot d\mathbf{P}, \quad (13.9)$$

$$F' = F_0 + V\mathbf{P} \cdot \mathbf{E}_0. \quad (13.10)$$

Sometimes free energy is used for the quantity

$$F'' = F' + \frac{E_0^2 V}{8\pi} = F_0 + \frac{E_0^2 V}{8\pi} + V\mathbf{P} \cdot \mathbf{E}_0, \quad (13.11)$$

which represents the free energy of the dielectric (allowing for its potential energy in the external field) plus the energy of the external electric field in the volume occupied by the dielectric.

Making use of (13.8), it can easily be shown that

$$\begin{aligned} dF'' &= dF_0 + V\mathbf{E} \cdot d\mathbf{P} + \frac{1}{4\pi} V\mathbf{E} \cdot d\mathbf{E} = \\ &= dF_0 + V\mathbf{E} \cdot \left(d\mathbf{P} + \frac{d\mathbf{E}}{4\pi} \right) = dF_0 + \frac{V\mathbf{E}}{4\pi} \cdot d\mathbf{D}. \end{aligned} \quad (13.12)$$

Thus, F'' represents the energy of the dielectric without field plus the energy of the field in the volume occupied by the dielectric.

It follows from (13.12) that

$$E_i = \frac{4\pi}{V} \frac{\partial F''}{\partial D_i}, \quad (13.13a)$$

$$\frac{\partial E_i}{\partial D_k} = \frac{4\pi}{V} \frac{\partial^2 F''}{\partial D_i \partial D_k} \quad (13.13b)$$

In an anisotropic medium the relation between \mathbf{D} and \mathbf{E} is given by formula (6.1) which we shall write in the form

$$E_i = \epsilon_{ik}^{-1} D_k,$$

where ϵ_{ik}^{-1} is the tensor inverse to ϵ_{ik} . Then (13.13) gives

$$\frac{V}{4\pi} \epsilon_{ik}^{-1} = \frac{\partial^2 F''}{\partial D_i \partial D_k} = \frac{\partial^2 F''}{\partial D_k \partial D_i} = \frac{V}{4\pi} \epsilon_{ki}^{-1}.$$

Thus ϵ_{ik} is a symmetric tensor.

For practical purposes it is convenient to use the thermodynamic potential instead of the free energy. Then instead of (13.8) we obtain

$$G = G_0 - \frac{1}{2}(\mathbf{P} \cdot \mathbf{E}_0) V. \quad (13.14)$$

The volume of the dielectric in the field is

$$V = \left(\frac{\partial G}{\partial p} \right)_T = V_0 - \frac{\mathbf{E}_0}{2} \cdot \left(\frac{\partial(\mathbf{P}V)}{\partial p} \right)_T.$$

The change in the volume of the dielectric when the external field is applied isothermally is called electrostriction. As can be seen from the last formula, the sign and magnitude of the effect depend on the polarizability of the body as well as on its compressibility.

When the field is applied, a change in the entropy of the dielectric occurs in addition to the change in the volume. Namely:

$$S = - \left(\frac{\partial G}{\partial T} \right)_p = S_0 + \frac{\mathbf{E}_0}{2} \cdot \left(\frac{\partial(\mathbf{P}V)}{\partial T} \right)_p. \quad (13.15)$$

The change in the entropy when the field is applied isothermally is accompanied by the release of heat $Q = T\Delta S$ (the so-called electro-caloric effect at constant pressure). When the field is applied adiabatically, a change in the temperature of the dielectric occurs.

As an example we shall find the electrostriction and electro-caloric effect for a sphere in a uniform field.

According to (12.9) we have

$$G = G_0 - \frac{1}{2}(\mathbf{P} \cdot \mathbf{E}_0) \quad V \approx G_0 - \frac{3}{8\pi} \frac{\epsilon_1 - 1}{\epsilon_1 + 2} V E_0^2,$$

whence

$$V = \left(\frac{\partial G}{\partial p} \right)_T = V_0 - \frac{3}{8\pi} \frac{\epsilon_1 - 1}{\epsilon_1 + 2} E_0^2 \left(\frac{\partial V}{\partial p} \right)_T - \frac{9E_0^2 V}{8\pi} \frac{1}{(\epsilon_1 + 2)^2} \left(\frac{\partial \epsilon_1}{\partial p} \right)_T. \quad (13.16)$$

Analogously

$$Q = T\Delta S = \frac{3TE_0^2}{8\pi} \frac{\epsilon_1 - 1}{\epsilon_2 + 2} \left(\frac{\partial V}{\partial T} \right)_p + \frac{9TE_0^2 V}{8\pi} \frac{1}{(\epsilon_1 + 2)^2} \left(\frac{\partial \epsilon_1}{\partial T} \right)_p. \quad (13.17)$$

It is easily shown that if ϵ decreases with increasing temperature (as is the case for most substances), $Q < 0$, i.e. as the polarization increases heat is released.

The next problem which naturally arises is the calculation of the electric susceptibility κ of the dielectric.

In contrast to the preceding relations which have a very general character and do not require the knowledge of the actual form of the partition function Z , to find the electric susceptibility it is necessary to obtain the partition function for the body.

Let us first consider the electric susceptibility of ideal gases. To find the value of the susceptibility per molecule it is necessary to find the corresponding partition function of the molecule in the electric field. One has to distinguish between two cases:

- (a) the molecules possess an intrinsic constant dipole moment,
- (b) the molecules possess no intrinsic dipole moment.

The dipole moment is defined by the expression

$$\mathbf{d} = \sum e_i \mathbf{r}_i,$$

where the summation is carried out over all charges in the molecule. In atoms and molecules of symmetric form positive and negative charges are distributed symmetrically. Because of this the summation over the negative and positive charges gives zero. All atoms and such symmetric molecules as H_2 , O_2 , CH_4 etc. have no intrinsic dipole moment. On the contrary, for strongly asym-

metric molecules, for example molecules made up of two different ions, such as HCl, HBr and so on, or molecules having an asymmetric form, such as CH_3Cl , CH_3COOH , H_2O , the dipole moment is different from zero. Such molecules are called polar molecules.

Let us first consider the properties of a gas with polar molecules having an intrinsic moment \mathbf{d}_0 . If $\epsilon^{(0)}$ is the energy of the molecule in the absence of a field, then, when it is placed in an electric field of strength \mathbf{E} , its energy will be equal to

$$\epsilon = \epsilon^{(0)} - \mathbf{d}_0 \cdot \mathbf{E} = \epsilon^{(0)} - d_0 |\mathbf{E}| \cos \theta, \quad (13.18)$$

where θ is the angle between the direction of the applied field and the axis of the molecule.

From formula (13.18) it follows that the change in the energy of a polar molecule in an electric field leads to the appearance of a potential energy equal to $(-d_0 |\mathbf{E}| \cos \theta)$ for the molecule. This energy in a uniform electric field does not depend on the position of the molecule and is determined solely by its orientation. The potential energy has a minimum value for a molecule oriented along the field, and a maximum value for a molecule oriented in the opposite direction. In the absence of a field, molecules are oriented completely at random; all orientations are equally probable. The electric field has an orienting effect and tends to arrange all the molecular dipoles along the field, where their potential energy is a minimum. The orientation of molecules along the field becomes more probable than against the field. The extent to which the field manages to orient all the dipoles is determined by the ratio of the energy $\mathbf{d}_0 \cdot \mathbf{E}$ acquired by the dipole molecules in the electric field to their thermal energy kT . If the latter is large, i.e. if $kT \gg \mathbf{d}_0 \cdot \mathbf{E}$, then the orienting effect of the field is relatively weak. On the contrary, if $kT \ll \mathbf{d}_0 \cdot \mathbf{E}$ all the dipoles will be oriented along the field.

The values of the dipole moments of the molecules of certain gases are given in table 1. In order of magnitude they are equal to the product of the charge of the electron and the size of the molecule.

Table 1

Gas	HCl	HBr	H ₂ O	SO ₂	CO ₂	CO	CHCl ₃	CH ₂ Cl ₂	CH ₃ Cl
Dipole moment (10^{-18} esu·cm)	1.03	0.79	1.84	1.61	0.00	0.12	0.95	1.59	1.89

By means of the data of table 1 the order of magnitude of $\mathbf{d}_0 \cdot \mathbf{E}/kT$ can be estimated. Because of the smallness of the dipole moment \mathbf{d}_0 it turns out that this quantity is very small for all temperatures at which the gases are still not condensed and in all practically attainable fields. In order that $\mathbf{d}_0 \cdot \mathbf{E}$ may be of the order of magnitude of kT it is necessary that $|\mathbf{E}|$ be of the order of magnitude of $kT/d_0 \approx 10^4 T$ V/cm. For $T \approx 300$ K we obtain $|\mathbf{E}| \approx 3 \times 10^6$ V/cm, which is obviously impractical. Thus, the orienting effect of the field is weak. Nevertheless, the appearance of a preferential orientation of molecular dipoles gives rise to a mean dipole moment \mathcal{P} of the entire gas different from zero.

We shall calculate the mean dipole moment of the gas according to formula (13.4). In this case, since the ratio $(\mathbf{d}_0 \cdot \mathbf{E})/kT$ is very small, the temperature can practically always be considered as high and the summation can be replaced by integration. The partition function of a gas of dipole molecules in an electric field has the form

$$Z = Z_0 \left(\int_0^\pi \exp \left(\frac{d_0 |\mathbf{E}| \cos \theta}{kT} \right) \sin \theta \, d\theta \int_0^{2\pi} d\varphi \right)^N. \quad (13.19)$$

Here N is the number of molecules in the gas, and Z_0 denotes the partition function of the gas in the absence of a field. The integration is carried out with respect to all states of the molecule, which are determined by the angles θ, φ . Since $d_0 |\mathbf{E}|/kT \ll 1$, we have

$$\begin{aligned} \int_0^\pi \exp \left(\frac{d_0 |\mathbf{E}| \cos \theta}{kT} \right) \sin \theta \, d\theta &\approx \\ &\approx \int_0^\pi \left\{ 1 + \frac{d_0 |\mathbf{E}|}{kT} \cos \theta + \left(\frac{d_0 |\mathbf{E}|}{kT} \right)^2 \frac{\cos^2 \theta}{2} \right\} \sin \theta \, d\theta = 2 + \frac{1}{3} \left(\frac{d_0 |\mathbf{E}|}{kT} \right)^2, \end{aligned}$$

so that

$$Z = Z_0 (4\pi)^N \left\{ 1 + \frac{1}{6} \left(\frac{d_0 |\mathbf{E}|}{kT} \right)^2 \right\}^N$$

On taking the logarithm we have

$$\ln Z \approx \ln Z_0 + N \ln 4\pi + \frac{N}{6} \left(\frac{d_0 |\mathbf{E}|}{kT} \right)^2. \quad (13.20)$$

Substituting (13.20) into (13.4), we find

$$\mathcal{P} = \frac{Nd_0^2|\mathbf{E}|}{3kT}. \quad (13.21)$$

This formula has a very simple meaning. If it is written in the form of the product of two factors

$$\mathcal{P} = \frac{d_0|\mathbf{E}|}{3kT} (d_0N),$$

then it is clear that the first factor characterizes the degree of orientation of the molecular dipoles, which is higher the stronger the applied field $|\mathbf{E}|$, and lower the stronger the disorienting effect of the thermal motion which is characterized by the value of kT . The second factor is the total dipole moment of the molecules in the gas. If all the dipoles were oriented along the field, the resulting polarization of the gas would be equal to d_0N . In the presence of thermal motion the polarization represents only a small fraction of this limiting polarization. One can picture the smallness of the orientation factor $d_0|\mathbf{E}|/kT$ in an obvious way as follows. Most of the polar molecules are moving and, in particular, are rotating in the gas with an energy which is considerably larger than the potential energy in the field d_0E . Hence the applied field cannot stop their rotation. As the molecules are rotating the dipole moment is averaged over all directions, and the mean contribution of a rotating molecule to the total dipole moment of the gas turns out to be equal to zero. However, there is in the gas a small percentage of molecules for which the kinetic energy of rotary motion is smaller than d_0E . Such molecules cannot be oriented against the field, and in the field their rotary motion is replaced by torsional oscillations about the direction of the field. It is only these molecules which contribute to the mean dipole moment of the gas.

From (13.21) one can find the electric susceptibility of a polar gas related to unit volume:

$$\kappa = \frac{\mathcal{P}}{|\mathbf{E}|V} = \frac{Nd_0^2}{3VkT}. \quad (13.22)$$

Formula (13.22) is called the Langevin formula, since it was first obtained by Langevin.

The electric susceptibility of the gas is connected with the directly measured dielectric constant by the relation (4.14):

$$\epsilon = 1 + 4\pi\kappa.$$

From the measurements of ϵ as a function of the temperature one can find the dipole moment d_0 of the molecules. The values of d_0 found in such a way are given in table 1. Their numerical value is small in comparison with unity, so that in gases ϵ is always of the order of unity.

The dipole moment is an important characteristic of molecules. In particular, it allows one to estimate the structure and geometrical form of the molecules. The more asymmetric the molecule, the larger is its dipole moment. This is seen, in particular, in the example of the molecules CCl_4 , CHCl_3 , CH_2Cl_2 , CH_3Cl and CH_4 , whose susceptibilities are shown in fig. IV.5. The value of d_0^2 is determined from the slope of the corresponding straight line.

Carbon tetrachloride and methane are symmetric molecules having the form of a tetrahedron; their dipole moments are nearly equal to zero. The slope of the corresponding straight line is connected with the induced dipole moment. For the chlorine derivatives of methane, which represent asymmetric molecules, the dipole moments differ from zero. The largest value occurs for the asymmetric molecule CH_3Cl .

The behaviour of molecules which possess no intrinsic dipole moment is of less interest. Under the action of an applied electric field they become polarized. The electron shells of the atoms or molecules are displaced with

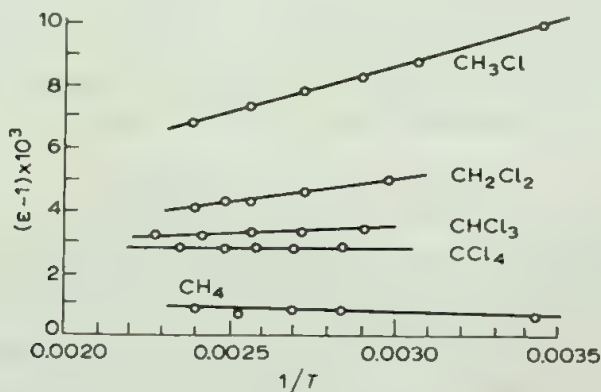


Fig. IV.5

respect to the nuclei, and an induced dipole moment arises in them. The value of the induced dipole moment is proportional to the strength of the applied field, so that

$$\mathbf{d} = \alpha \mathbf{E}$$

and the energy is

$$\epsilon = \epsilon^{(0)} - \frac{1}{2} \alpha |\mathbf{E}|^2. \quad (13.23)$$

Hence

$$Z = Z_0 \exp\left(\frac{N\alpha |\mathbf{E}|^2}{2kT}\right)$$

and

$$P = N\alpha |\mathbf{E}|. \quad (13.24)$$

The polarization turns out to be independent of the temperature. In other words, the thermal motion has no effect on the polarization of the electron shells of atoms and molecules.

The theory given here is valid only for very rarefied gases. For more dense gases and particularly for liquids the interaction between molecules plays an essential role in their electric properties. In such systems one can no longer consider the orienting effect of the field on an individual dipole. The orientation of the dipoles will be determined not only by the external but also by the internal electric field which is produced by all the polar molecules.

The theory of polar liquids turns out to be very complex, and we cannot develop it here. We note only that in this field things are not as clear as in the case of rarefied gases.

The dielectric polarization of most solid dielectrics is due to the induced polarization of the molecules of the dielectric. Even in crystals made up of polar molecules no orientation polarization occurs, since the molecules in the crystal interact strongly with their neighbours and the electric field is too weak to overcome the forces of interaction and to turn the molecules. One of a few typical exceptions is the solid HCl. The measurements of I.V. Kurchatov have shown that the molecules of HCl are turned under the action of an electric field and an orientation polarization takes place in the crystal. The values of the susceptibility and dielectric constant for solid dielectrics are

much larger than for gases. The dielectric constants of solid dielectrics have values several times unity, and in a number of cases reach as much as 100.

In view of the large value of the dielectric constant in a dense medium it is necessary to take into account that a molecule in the medium is acted upon not by the external field E_0 but by the effective internal field made up of the external field and the field which arises in the medium because of the polarization of its molecules.

The dielectric constant of solid dielectrics varies relatively weakly with the temperature. This means that the dielectric properties of solid bodies are associated with the change of the charge distribution inside the molecules but do not depend on their thermal motion.

Direct Electric Current and the Magnetic Properties of Matter

§ 14. Ohm's law

Having considered the basic properties of the electrostatic field of charges at rest, we can pass on to the study of the more complex case of the electromagnetic field arising when there is a steady motion of free charges, i.e. in the presence of a current in conductors constant in time. This problem has two aspects which are to a considerable degree independent:

- (1) to find the electromagnetic field of direct currents,
- (2) the consideration of the mechanism of passage of the current in different media, i.e. the mechanism of electrical conductivity.

Without going into a study of the mechanism of electrical conductivity, we shall restrict ourselves to the assumption that the current density in a homogeneous conductor is related to the field strength by Ohm's law ($\mathbf{j} = \sigma \mathbf{E}$).

The value of the electrical conductivity σ is closely associated with the mechanism of the passage of the current and varies over a very wide range for different conductors.

In Part V we shall consider the microscopic meaning of the electrical conductivity and shall estimate its value for some important conductors.

We shall further assume that the constancy of the electric current in time is maintained by devices called current sources. Examples of such sources are

constant-current generators of various types, i.e. galvanic cells, accumulators, thermocouples etc. Indeed, it is physically obvious that no combination of charged or neutral conductors can ensure the passage of a constant current in a system. By bringing conductors at different potentials into contact we can cause a transient motion of free charges which will continue until the potentials of all the conductors become the same.

The sources of a constant electric current must always have a non-electrostatic character. They can, for example, have an electrochemical character. In what follows we shall not go into details of the current sources. For us, only the fact that the current sources ensure the maintainance of a constant current is of importance.

Formally, without considering the mechanism of the current sources, we can include them in the composition of the system of conductors considered by changing the form of Ohm's law. That is, observing that the current sources produce a current in the conductors independent of the direct action of the electric field, we write Ohm's law in the generalized form

$$\mathbf{j} = \sigma(\mathbf{E} + \mathbf{E}^{\text{imp}}). \quad (14.1)$$

The vector \mathbf{E}^{imp} (where imp stands for "impressed"), which depends on the coordinates, formally characterizes the action of the current sources. In those parts of the conductor where the current source acts, for $\mathbf{E} = 0$, $\mathbf{j} \neq 0$. Thus, for example, if the system is made up of a conductor connected to the plates of a galvanic cell, then in the region of space occupied by the cell $\mathbf{E}^{\text{imp}} \neq 0$. The processes occurring inside the galvanic cell make it possible to maintain its plates, which are connected to the conductor, at different potentials. This in its turn ensures the existence of an electric field constant in time inside the cell and, in correspondence with Ohm's law, the existence of a constant current in the conductor.

The quantity \mathbf{E}^{imp} is for historical reasons called, not quite correctly, the impressed electromotive force (\mathbf{E}^{imp} does not have the dimensions of a force and is not, in essence, the analogue of a force). The electromotive force is a quantitative characteristic of the device which maintains the passage of a constant current in conductors. We shall return to the interpretation of the notation of the electromotive force in later sections.

Maxwell's equations for a constant current in the presence of electromotive forces has the form

$$\left. \begin{aligned} \nabla \times \mathbf{E} &= 0, \\ \nabla \cdot \mathbf{D} &= 4\pi\rho; \end{aligned} \right\} \quad (14.2)$$

$$\left. \begin{aligned} \nabla \times \mathbf{H} &= \frac{4\pi}{c} \mathbf{j}, \\ \nabla \cdot \mathbf{B} &= 0. \end{aligned} \right| \quad (14.3)$$

The equation of continuity for the passage of steady currents can be written in the form

$$\nabla \cdot \mathbf{j} = 0. \quad (14.4)$$

It is easily seen that the system of equations written above is complete. We see, first of all, that the electric field distribution does not depend on the magnetic field distribution. The latter is determined by defining the current density \mathbf{j} over all space.

At the interface of conducting media the following boundary conditions must be fulfilled:

$$E_{\text{tg}}^{(1)} = E_{\text{tg}}^{(2)}, \quad (14.5)$$

or

$$\frac{j_{\text{tg}}^{(1)}}{\sigma^{(1)}} = \frac{j_{\text{tg}}^{(2)}}{\sigma^{(2)}} \quad (14.6)$$

and

$$j_n^{(1)} = j_n^{(2)}. \quad (14.7)$$

The first of these is the same as (5.2), and the second is obtained from the equation of continuity in the same way as, for example, the condition (5.5).

The boundary conditions (14.6) – (14.7) can be interpreted in an obvious way as the refraction of lines of flow at the interface according to the law

$$\frac{\text{tg } \alpha_1}{\text{tg } \alpha_2} = \frac{\sigma^{(1)}}{\sigma^{(2)}}, \quad (14.8)$$

where α is the angle between the line of flow and the normal to the surface in the corresponding media.

At the interface conductor–dielectric the following boundary condition is satisfied:

$$j_n = 0. \quad (14.9)$$

Magnetic vectors satisfy the boundary conditions which have been considered in § 5.

§ 15. A linear conductor carrying a constant current

We shall first consider the very important case of a linear conductor with constant current. By a linear conductor we shall mean one whose length is very large compared with its transverse dimensions. Linear conductors are also often called wires. The current density vector in a linear conductor can, by virtue of the boundary condition (14.9) at its surface, be considered with a high degree of accuracy to be parallel to the vector $d\mathbf{l}$ which is tangential to the axis of the conductor. Thus, for each point of a linear conductor one can write

$$\mathbf{j} d\mathbf{l} = j d\mathbf{l}. \quad (15.1)$$

We bring into consideration the total current I flowing through the cross-section of the linear conductor normal to the axis of the conductor (or, what is the same, normal to the flow lines). By definition

$$I = \int \mathbf{j} \cdot d\mathbf{S} = \int \sigma(\mathbf{E} + \mathbf{E}^{\text{imp}}) \cdot d\mathbf{S}, \quad (15.2)$$

where the integration is carried out over a cross-section of the contour carrying the current.

From the equation of continuity (14.4) and the boundary condition one can write

$$\oint \mathbf{j} \cdot d\mathbf{S} = 0,$$

or

$$I = \int \mathbf{j} \cdot d\mathbf{S} = jS = \text{const}, \quad (15.3)$$

where S is the cross-section of the conductor at a given position.

The equation of continuity in integral form shows that the same current I flows through any section of the linear conductor.

We integrate the formula of the generalized Ohm's law (14.1) along the conductor carrying the current. We have

$$\int_1^2 \mathbf{j} \cdot \frac{d\mathbf{l}}{\sigma} = \int_1^2 \mathbf{E} \cdot d\mathbf{l} + \int_1^2 \mathbf{E}^{\text{imp}} \cdot d\mathbf{l}.$$

We transform the first integral, writing

$$\int_1^2 \mathbf{j} \cdot \frac{d\mathbf{l}}{\sigma} = \int_1^2 j \frac{dS}{\sigma S} = I \int_1^2 \frac{dl}{\sigma S} = IR_{12}, \quad (15.4)$$

where R_{12} represents the ohmic resistance of the conductor along the segment (1, 2). Then we have

$$IR_{12} = \varphi_1 - \varphi_2 + \mathcal{E}_{12}^{\text{imp}}, \quad (15.5)$$

where $\varphi_1 - \varphi_2$ is the potential difference between the points 1 and 2, and

$$\mathcal{E}_{12}^{\text{imp}} = \int_1^2 \mathbf{E}^{\text{imp}} \cdot d\mathbf{l} \quad (15.6)$$

is called the impressed electromotive force (e.m.f.) over the segment (1, 2).

If there is no electromotive force in a given segment of the conductor, i.e. $\mathbf{E}^{\text{imp}} = 0$ then (15.5) goes over into the simple Ohm's law.

If the path is closed and the points 1 and 2 coincide, since by virtue of (14.2) the field \mathbf{E} has the character of potential, the integral $\oint \mathbf{E} \cdot d\mathbf{l} = 0$ and

$$IR = \mathcal{E}^{\text{imp}}, \quad (15.7)$$

where R is the resistance of the entire linear contour and

$$\mathcal{E}^{\text{imp}} = \oint \mathbf{E}^{\text{imp}} \cdot d\mathbf{l}. \quad (15.8)$$

The product of the current and the resistance of the current carrying filament is equal to the e.m.f. round the closed circuit.

Let us now consider the energy relations for a current carrying filament.

As we have stressed, the entire work done by the current in a direct current circuit goes into heat. Hence the total heat released in the linear conductor is

$$\begin{aligned}
 Q &= \frac{j^2}{\sigma} dV = \int j \cdot (E + E^{\text{imp}}) dV = \\
 &= - \int j \cdot \nabla \varphi dV + \int j \cdot E^{\text{imp}} dV = \\
 &= \int j \cdot E^{\text{imp}} dV - \int (\nabla \cdot j \varphi) dV + \int \varphi (\nabla \cdot j) dV = \\
 &= \int j \cdot E^{\text{imp}} dV - \oint \varphi j_n dS = \int j \cdot E^{\text{imp}} dV
 \end{aligned}$$

by virtue of (14.2), (14.4) and (14.9).

The total heat released in the circuit turns out to be equal to the work done by the electromotive forces.

§ 16. Direct current in a conducting medium

Another limiting case is the passage of current in a system of good conductors (for example, metal electrodes) immersed into a conducting medium. If it is assumed that electromotive forces are absent in the conducting medium, the equations for the electric field can be written in the form

$$\nabla \times E = 0,$$

$$\nabla \cdot j = \nabla \cdot \sigma E = 0.$$

In a homogeneous medium, for $\sigma = \text{const}$, the last expression assumes the form

$$\nabla \cdot E = 0.$$

Introducing the field potential φ , we find that it satisfies the equation

$$\nabla^2 \varphi = 0. \quad (16.1)$$

At the surface of the conductors the boundary conditions (14.5) and (14.7) are satisfied. They can be written in terms of the potential φ . The condition (14.5) for the continuity of the tangential component of the field goes over directly into the condition of the continuity of the potential at the interface:

$$\varphi_1 = \varphi_2 . \quad (16.2)$$

The equality of the normal components of the current density $j_n = \sigma E_n$ gives

$$\sigma^{(1)} \left(\frac{\partial \varphi}{\partial n} \right)_1 = \sigma^{(2)} \left(\frac{\partial \varphi}{\partial n} \right)_2 , \quad (16.3)$$

where n is the normal to the interface.

We see that the equation for the potential and the whole set of boundary conditions, which determine the current distribution in a conducting medium, are identical with the corresponding expressions of §8 which determine the electrostatic field distribution in two dielectric media. The only difference lies in the fact that the electric conductivities $\sigma^{(1)}$ and $\sigma^{(2)}$ of the media appear in the boundary condition (16.3) instead of the dielectric constants ϵ_1 and ϵ_2 .

Hence the potential in a conducting medium is determined by the formulae of electrostatics with the substitution of σ for ϵ .

Let us consider the case of two electrodes in an infinite medium. We write the current flowing from an electrode in the form

$$I = \oint \mathbf{j} \cdot d\mathbf{S} = \oint j_n dS = \sigma \oint \left(\frac{\partial \varphi}{\partial n} \right)_S dS .$$

Introducing the electrostatic capacitance according to formula (8.13), we find

$$I = \frac{4\pi\sigma}{\epsilon_1} \left[\frac{\epsilon_1}{4\pi} \oint \left(\frac{\partial \varphi}{\partial n} \right)_S dS \right] = \frac{4\pi\sigma}{\epsilon_1} C\varphi_1 .$$

The resistance is equal to

$$R = \frac{\varphi_1 - \varphi_2}{I} = \frac{\varphi_1}{I} = \frac{\epsilon_1}{4\pi\sigma} \frac{1}{C} . \quad (16.4)$$

This last formula allows one to express the resistance of the system formally in terms of the capacities of an electrostatic system of conductors with analogous geometric characteristics.

As an example a system of two spherical electrodes immersed into an infinite medium can be considered. We assume that the radii of the electrodes a and b are small in comparison to the distance between their centres.

The solution of Laplace's equation for the two spheres has the form

$$\varphi = \frac{a}{r_1} \varphi_a - \frac{b}{r_2} \varphi_b,$$

where φ_a and φ_b are the potentials at the surfaces of the spheres, and r_1 and r_2 are the distances to the centres of the spheres.

The total current from the surface of the first sphere is equal to

$$I = \oint \sigma \left(\frac{\partial \varphi}{\partial r} \right)_{r=a} dS \approx 4\pi a \sigma \varphi_a,$$

if the second term of φ , which for $r = a$ is small in comparison with the first term, is disregarded. Analogously, the total current to the second sphere turns out to be equal to

$$I = -4\pi b \sigma \varphi_b.$$

The resistance is

$$R = \frac{\varphi_a - \varphi_b}{I} \approx \frac{1}{4\pi\sigma} \left(\frac{1}{a} + \frac{1}{b} \right).$$

Thus the solution of the problem of the spatial current distribution reduces to the solution of the corresponding electrostatic problem.

However, an important reservation should be made. If the conductor is partially contiguous to a conducting medium and partially to a non-conducting medium, the formal analogy with electrostatics is meaningless. Indeed, in a non-conducting medium $\sigma = 0$, whereas in electrostatics there are no bodies with a dielectric constant equal to zero.

We shall find one more useful relation between the current density and total current in a conductor in the general case. Let the current density flowing along a conductor be distributed non-uniformly over the cross-section. We divide the conductor into arbitrarily fine current tubes. The solenoidal character of direct current always allows one to make such a division. Each current tube can be considered a linear conductor, and we can write

$$dI = \mathbf{j}_\alpha \cdot d\mathbf{S}_\alpha, \quad (16.5)$$

where the index α denotes the ordinal number of the tube.

This equation has a simple meaning: although the current density distribution is determined by the physical properties of the conductor and by its geometry, the current density is proportional to the total current, other things being equal.

If the current density were distributed uniformly over the cross-section, one could obviously write

$$\frac{dI_\alpha}{I} = \frac{dS_\alpha}{S}, \quad (16.6)$$

where I and S are the total current and the cross-section of the conductor respectively. For a non-uniform current distribution one can always assume that

$$\frac{dI_\alpha}{I} = \frac{dS_\alpha}{S} \psi = \text{const} \quad (\text{along the length of the conductor}), \quad (16.7)$$

where the function ψ characterizes the non-uniformity of the current distribution. Hence for any current distribution we have

$$\mathbf{j} = I \frac{\psi}{S} \mathbf{l}, \quad (16.8)$$

where \mathbf{l} is the unit vector directed along the flow line.

§ 17. The magnetic fields of direct currents. The Biot—Savart law

Knowing the current density distribution one can find the magnetic field distribution by integrating eqs. (14.3). Introducing the vector potential A into (14.3), according to formula (4.19) we obtain the equation

$$\nabla \times \left(\frac{1}{\mu} \nabla \times A \right) = \frac{4\pi}{c} \mathbf{j}. \quad (17.1)$$

For an infinite homogeneous and isotropic medium characterized by a constant magnetic permeability μ one can write

$$\nabla^2 A = - \frac{4\pi\mu}{c} \mathbf{j}. \quad (17.2)$$

The solution of this last equation

$$\mathbf{A} = \frac{\mu}{c} \int \mathbf{j} \frac{dV}{r} \quad (17.3)$$

differs only by the factor μ from the solution of eq. (19.13) of Part I for the vector potential in vacuum.

Since outside the conductors the current density \mathbf{j} is equal to zero, the integration can actually be carried out only with respect to the volume of the conductors. However, we assume that μ has one and the same value both in the conductors and in the medium which surrounds them. If the system contains no ferromagnetic bodies, then the actual value of μ differs little from unity. Hence μ can be assumed to have approximately the same value over all space. In the presence of ferromagnetics the formula (17.3) becomes meaningless.

For linear conductors one can essentially simplify formula (17.3), writing, by virtue of the constancy of the total current I over the cross-section of the linear conductor,

$$\mathbf{A} = \frac{\mu}{c} \int \mathbf{j} \frac{dV}{r} = \frac{\mu}{c} \int \frac{j dS dl}{r} = \frac{\mu I}{c} \int \frac{dl}{r}. \quad (17.4)$$

In formula (17.4) μ denotes the magnetic permeability of the medium, which is external to the conductor carrying the current. We see that all characteristics of the linear conductor itself, for example the current density distribution in it, are eliminated in (17.4), and the conductor is considered as a purely geometric object. Hence the properties of the linear conductor, including also the magnetic properties of the material of which the wire is made, do not affect the value of the magnetic field. Formula (17.4) is valid for all conductors, including ferromagnetic ones.

From the definition of the vector potential it follows that the magnetic induction of a linear conductor carrying a current I is equal to (see (19.15) of Part I)

$$\begin{aligned} \mathbf{B} &= \nabla \times \mathbf{A} = \frac{\mu I}{c} \nabla \times \int \frac{dl}{r} = \frac{\mu I}{c} \int \nabla \times \frac{dl}{r} = \\ &= \frac{\mu I}{c} \int (\nabla r^{-1}) \times dl = \frac{\mu I}{c} \int \frac{dl \times \mathbf{r}}{r^3}. \end{aligned} \quad (17.5)$$

Formula (17.5) expresses the Biot-Savart law for a homogeneous and isotropic medium. The induction (mean field) \mathbf{B} turns out to be larger by a factor of μ than the field of the same current in vacuum. Correspondingly the field strength in the medium $\mathbf{H} = \mu^{-1}\mathbf{B}$ is the same as the field in a vacuum.

The Biot-Savart law is often written in differential form

$$d\mathbf{B} = \frac{\mu I}{c} \frac{d\mathbf{l} \times \mathbf{r}}{r^3}, \quad (17.6)$$

where $d\mathbf{B}$ is the contribution to the induction of the current element $d\mathbf{l}$. It should be kept in mind that (17.5) cannot be resolved unambiguously into the elements (17.6). One can always add to (17.6) a vector function which reduces to zero in integrating over a closed contour carrying a direct current.

As an example of the application of the formulae obtained, we shall calculate the magnetic field produced in the surrounding space by a current flowing in an infinite straight linear conductor.

From considerations of symmetry it is clear that the field is directed along tangents to circles which are concentric about the conductor. The Biot-Savart law applied to the case considered gives

$$\mathbf{B} = \frac{\mu I}{c} \int \frac{d\mathbf{l} \times \mathbf{n}}{r^2},$$

where $\mathbf{n} = \mathbf{r}/r$. Introducing the shortest distance ρ from the point of observation to the conductor, we have for the field component B_ψ directed along a tangent to the circle of radius ρ concentric with respect to the conductor

$$B_\psi = \frac{\mu I}{c} \int \frac{\cos^2 \alpha}{\rho^2} \sin(90^\circ - \alpha) dl = \frac{\mu I}{c\rho} \int_{-\frac{1}{2}\pi}^{+\frac{1}{2}\pi} \cos \alpha d\alpha = \frac{2\mu I}{c\rho}, \quad (17.7)$$

where we have set (fig. IV.6)

$$r \cos \alpha = \rho,$$

$$\sin(90^\circ - \alpha) dl = r d\alpha = \frac{\rho d\alpha}{\cos \alpha}.$$

Further, we calculate the magnetic field on the central axis, z , perpendicular to the plane of a flat circular current carrying conductor of radius a with centre on the axis z (fig. IV.7). In this case one can write for the field component

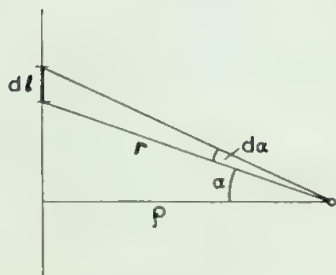


Fig. IV.6

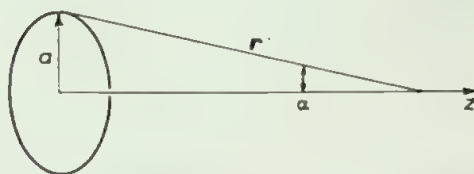


Fig. IV.7

$$B_z = \frac{\mu I}{c} \oint \frac{[d\mathbf{l} \times \mathbf{r}]_z}{r^3} = \frac{\mu I}{c} \int \frac{dl \sin \alpha}{r^2} = \frac{\mu I \cos^2 \alpha \sin \alpha}{cz^2} 2\pi a.$$

Since (fig. IV.7)

$$\cos \alpha = z(z^2 + a^2)^{-\frac{1}{2}}, \quad \sin \alpha = a(z^2 + a^2)^{-\frac{1}{2}},$$

we find finally

$$B_z = \frac{2\mu I}{c} \frac{S}{(z^2 + a^2)^{\frac{3}{2}}}, \quad (17.8)$$

where S is the area of the circle formed by the current.

In particular, at a large distance from the conductor ($z \gg a$)

$$B_z = \frac{2\mu IS}{cz^3} = \frac{2\mu M_z}{z^3}, \quad (17.9)$$

where the projection M_z of the magnetic moment of the current is equal to IS/c (see §22 of Part I).

At the centre of the circle in the plane $z = 0$

$$B_z = \frac{2\mu I}{ca^3}. \quad (17.10)$$

The field component B_{\perp} perpendicular to the z -axis is equal to zero, as is clear from considerations of symmetry: to opposite segments of the current carrying conductor there will correspond values of B_{\perp} with opposite signs.

Finding the magnetic field in non-linear conductors carrying direct currents presents a problem which is complex from the mathematical standpoint. The calculation of the vector potential according to the general formula (17.3) can be carried out only for individual cases. We shall therefore confine ourselves to examples.

A very simple example is the magnetic field of an infinite current carrying linear conductor of circular cross-section of radius a . In this case the integration of eq. (14.3) can be carried out directly. Owing to the cylindrical symmetry the magnetic field inside as well as outside the conductor has only one component, H_ψ . The integration of (14.3) over the area of a circle of a radius $\rho < a$ gives

$$2\pi\rho H_\psi = \frac{4\pi}{c} jS,$$

where $S = \pi\rho^2$. This last formula can be written in the form

$$H_\psi = \frac{2I}{c} \frac{\rho}{a^2} \quad (\rho < a), \quad (17.11)$$

or, in the vector form,

$$\mathbf{H} = \frac{2I}{ca^2} [\mathbf{l} \times \boldsymbol{\rho}] = \frac{2\pi j}{c} [\mathbf{l} \times \boldsymbol{\rho}], \quad (17.12)$$

where \mathbf{l} is a unit vector along the axis of the cylinder. Analogously, outside the conductor

$$H_\psi = \frac{2I}{c\rho} \quad (\rho > a). \quad (17.13)$$

If the cross-section of the conductor is more complex, then the principle of superposition of fields sometimes proves useful.

§18. The magnetization of magnetic materials and the magnetic moment

We now pass on to the consideration of the magnetic properties of matter.

If a system of particles is placed in an external magnetic field, then the system will be magnetized. The mean magnetic moment of the system can be determined by the same relation as the mean dipole moment:

$$\mathbf{M} = kT \frac{\partial}{\partial \mathbf{H}} \ln Z = - \frac{\partial F}{\partial \mathbf{H}}, \quad (18.1)$$

where \mathbf{H} is the strength of the external magnetic field.

Correspondingly one can write the expression for the change in the free energy:

$$dF = dF_0 - \mathbf{M}V \cdot d\mathbf{H}.$$

One can also introduce the Gibbs thermodynamic potential

$$dG = dG_0 - \mathbf{M}V \cdot d\mathbf{H} \quad (18.2)$$

Their meanings are analogous to those of the corresponding quantities for the electric field.

We shall not dwell on the analysis of the phenomena of magnetostriction and the magnetocaloric effect. They are also analogous to the electric effects which have been considered in § 13.

The magnetic susceptibility can be calculated according to the formula

$$\chi = \frac{|\mathbf{M}|}{|\mathbf{H}|} = \frac{kT}{V} \frac{1}{|\mathbf{H}|} \frac{\partial \ln Z}{\partial H}. \quad (18.3)$$

Depending on the sign of χ one distinguishes between diamagnetic ($\chi < 0$) and paramagnetic ($\chi > 0$) substances. In addition to diamagnetic and paramagnetic substances there is a special group of ferromagnetic bodies, for which the magnetic susceptibility is extremely large and depends strongly on the magnetic field.

For diamagnetic substances the magnetic susceptibility is usually very small in absolute value ($\chi \approx 10^{-6}$) per gram-mole and does not depend on the temperature. All inert gases, many gases whose molecules represent saturated chemical compounds, almost all organic compounds, all simple insulators and about one half of all metals (Cu, Ag, Au, Hg, Zn for instance) are diamagnetic. Among the latter one encounters the so-called anomalous diamagnetics, for which the susceptibility exceeds the normal value quoted above by a factor of 10–100 and which have a number of other anomalous properties (for example, χ depends on the temperature and on the field, as in the case of Bi and Sb).

For normal paramagnetics the magnetic susceptibility depends on the temperature according to the law

$$\chi = \frac{\text{const}}{T}.$$

In order of magnitude χ is about 10^{-4} – 10^{-6} per gram-mole. Among such paramagnetic substances there are certain gases (O_2 , NO , CO_2 for instance), the crystalline hydrates of rare earth salts (for example, $\text{Gd}_2\text{SO}_4 \cdot 8\text{H}_2\text{O}$), the salts of the metals of the group of platinum, iron etc. For many normal paramagnetics the dependence of χ on the temperature has the form

$$\chi = \frac{\text{const}}{T \pm \Delta},$$

where Δ is a constant. Another group of paramagnetic substances is made up of paramagnetic metals possessing a small paramagnetic susceptibility ($\chi = 10^{-6}$ – 10^{-7} per gram-mole) which does not depend on the temperature. There also exist the so-called anomalous paramagnetics, for which the paramagnetic susceptibility depends on the field (metamagnetics) or has a maximum value at a certain temperature (antiferromagnetics).

Finally, substances with a very large positive susceptibility (in order of magnitude reaching 10^3) which depends in a complex way on the magnetic field strength and on the temperature and a number of other factors make up the group of ferromagnetics.

We cannot here discuss in detail all aspects of the modern theory of the magnetic properties of matter. We shall confine ourselves only to certain general observations and to the exposition of the theory of the paramagnetism of normal paramagnetics. The theory of the diamagnetic properties of atoms will be substantiated in ch. 9 of Part V.

Before passing on to the exposition of the modern theory of the magnetic properties of matter it is necessary to dwell briefly on a statement which seems at first sight to be paradoxical. Namely, one can in the most general form prove that the magnetic moment of any body calculated by means of the laws of classical statistics is identically equal to zero. We shall present the simplest proof of this theorem.

Any system in an external field can be looked on as a set of moving charged particles. As is known from electrodynamics, when a charged particle is moving in a uniform magnetic field directed along the z -axis the generalized momentum of the particle has the form (§41 of Part I)

$$p_x = p_x^{(0)} - \frac{eHy}{2c}, \quad p_y = p_y^{(0)} + \frac{eHx}{2c}, \quad p_z = p_z^{(0)},$$

where $p_x^{(0)}$, $p_y^{(0)}$, $p_z^{(0)}$ are the components of the momentum in the absence of field.

The Hamilton function is given by the expression (41.8) of Part I:

$$H = \frac{[p_x + (eHy/2c)]^2 + [p_y - (eHx/2c)]^2 + p_z^2}{2m}.$$

The partition function of the system in the magnetic field has the form

$$Z = \left\{ \int e^{-H/kT} \frac{d\mathbf{p} dV}{h^3} \right\}^N.$$

Introducing instead of p_x a new variable

$$p'_x = p_x^{(0)} = p_x + \frac{eHy}{2c},$$

we see that upon integration with respect to p'_x from minus infinity to plus infinity the partition function turns out to be independent of the external field \mathbf{H} . The same holds also for p_y . By virtue of (18.3) the mean magnetic moment is identically equal to zero.

This seems to be particularly paradoxical because most books give the classical interpretation of diamagnetism and paramagnetism. The diamagnetic properties of matter are associated with the change in the orbital motion of electrons in the atom caused by the magnetic field. As is well known, the magnetic field induces a current flowing in a closed path in such a direction that the additional magnetic field of the current weakens the applied field. The induced magnetic moment of the current is directed against the field and is proportional to the strength of the latter as well as to the area embraced by the contour. Assuming the electron moving in the atom to be a current filament, one can obtain the following purely electrodynamic expression for the magnetic moment of the particle moving in an orbit of radius r_0 :

$$\mu = -\frac{e^2}{6mc^2} r_0^2 H.$$

As to paramagnetism, it is associated in classical electrodynamics with the presence of a magnetic moment of the electron moving in orbit and having an angular momentum different from zero. If L denotes the angular momentum of the system, then in classical electrodynamics it turns out that the system possesses the magnetic moment

$$\mu = \frac{e}{2mc} L. \quad (18.4)$$

The magnetic moment defined by formula (18.4) is the analogue of the electric dipole moment. Every atom is a sort of a small magnet. Hence the above considerations and formula (18.1) are completely applicable to it. In an atomic gas whose atoms possess magnetic moment μ there must arise a mean magnetic moment owing to the appearance of the preferential orientation of magnetic moments along the field.

The inconsistency of these arguments lies in the fact that one assumes beforehand the existence of stable electron orbits. However, it is well known that the existence of stable orbits cannot be conceived on the basis of classical concepts. The absence of a magnetic moment in classical physics is an expression of the fact of the absence of stable motion in a system of elementary charges. The assumption of the existence of stationary electron orbits in atoms or fixed moments of atoms, which is used in the "classical" theory of magnetism, represents in essence the assumption of the quantization of states, in an implicit form.

It is proved in quantum mechanics that the energy levels of an atomic system change in a magnetic field. In the case of atoms or ions, for which the mean value of the angular momentum L of the system is equal to zero, for the energy of the ground state one obtains the following expression (see ch. 9 of Part V):

$$\epsilon = \epsilon_0 + \frac{e^2 H^2}{12mc^2} \sum (r_i^2)_{\text{mean}}, \quad (18.5)$$

where ϵ_0 is the value of the energy in the ground state, and $(r_i^2)_{\text{mean}}$ is the quantum-mechanical mean of the radius vector of the i th electron in the ground state. The summation is carried out over all electrons in the atom. To this energy there corresponds a mean magnetic moment

$$\mu_{\text{mean}} = -\frac{\partial \epsilon}{\partial H} = -\frac{e^2 H}{6mc^2} \sum (r_i^2)_{\text{mean}}. \quad (18.6)$$

Formula (18.6) is in form the same as the classical formula given above, but has a different meaning: the quantity $(r^2)_{\text{mean}}$ represents the quantum-mechanical mean (see ch. 3 of Part V).

The atom or ion in the ground state has a diamagnetic susceptibility

$$\chi_0 = -\frac{e^2}{6mc^2} \sum (r_i^2)_{\text{mean}}. \quad (18.7)$$

We stress that in (18.7) no averaging over different states is carried out, and the averaging has nothing in common with statistical averaging.

A system representing a set of N independent atoms or ions will, in the ground state, possess an induced magnetic moment

$$M = -\frac{Ne^2H}{6mc^2} \sum (r_i^2)_{\text{mean}}$$

and a diamagnetic susceptibility

$$\chi = -\frac{Ne^2}{6mc^2} \sum (r_i^2)_{\text{mean}}. \quad (18.8)$$

If one substitutes values for $(r^2)_{\text{mean}}$ calculated by different methods (for these methods see Part V) or the mean radii of atoms obtained from the kinetic theory of gases, then one obtains for χ a value which agrees with those found from measurements.

If a system (an atom or an ion) in the ground state possesses an angular momentum different from zero, then the energy in a magnetic field will have a different form.

It turns out that in quantum mechanics, if a molecular system possesses an orbital angular momentum L , then its energy in a state i is equal to

$$\epsilon_i = \epsilon_i^{(0)} - \frac{eh}{4\pi mc} L_z H + \frac{e^2 H^2}{6mc^2} \sum (r_i^2)_{\text{mean}}, \quad (18.9)$$

where L_z is the projection of the angular momentum onto the direction of the magnetic field (the z -axis is chosen to be along the field). For the derivation of (18.9) see ch. 9 of Part V.

Simple estimates show that the last term of formula (18.9) for all values of the field strength H is small in comparison with the second term. Exceptions are for very large organic molecules, for which $\sum (r_i^2)_{\text{mean}}$ is very large.

In what follows we shall drop the last term of (18.9) and write the energy in the form

$$\epsilon_i = \epsilon_i^{(0)} - \frac{eh}{4\pi mc} L_z H. \quad (18.10)$$

It is shown in quantum mechanics (see ch. 3 of Part V) that the projection of the angular momentum onto the z -axis takes on discrete values:

$$L_z = -L, \quad -L+1, \dots, 0, \quad 1, \dots, L-1, \quad L$$

(all together $2L+1$ values). Hence in a magnetic field the i th energy level splits into $2L+1$ levels which possess the energies

$$\begin{aligned} \epsilon_i &= \epsilon_i^{(0)} + \frac{eh}{4\pi mc} LH; \\ \epsilon_i^{(0)} + \frac{eh}{4\pi mc} (L-1)H; \\ &\dots \dots \dots; \\ \epsilon_i^{(0)} - \frac{eh}{4\pi mc} (L-1)H; \\ \epsilon_i^{(0)} - \frac{eh}{4\pi mc} LH. \end{aligned}$$

The presence of an intrinsic angular momentum (spin) as well as the angular momentum due to the orbital motion in a system of electrons leads to the appearance of a spin magnetic moment in the system. If the resulting spin S of the system is different from zero, while the resulting orbital momentum L is equal to zero, the energy of the system in a magnetic field turns out to be equal to (see ch. 8 of Part V)

$$\epsilon_i = \epsilon_i^{(0)} - \frac{eh}{2\pi mc} S_z H, \quad (18.11)$$

where S_z is the projection of the intrinsic angular momentum onto the direction of the field taking on a discrete series of values:

$$S_z = -S, \quad -S+1, \dots, S-1, \quad S.$$

For such a system the i th level splits into $2S+1$ sublevels:

$$\epsilon_i = \epsilon_i^{(0)} + \frac{eh}{2\pi mc} SH, \dots, \epsilon_i^{(0)} - \frac{eh}{2\pi mc} SH. \quad (18.12)$$

We shall not consider the general case where L and S simultaneously differ from zero.

In a magnetic field the system in the i th state will possess a mean (in the quantum-mechanical sense) magnetic moment

$$(\mu_i)_{\text{mean}} = - \frac{\partial \epsilon_i}{\partial H} = \frac{eh}{4\pi mc} L_z, \quad (18.13)$$

or

$$(\mu_i)_{\text{mean}} = - \frac{\partial \epsilon_i}{\partial H} = \frac{eh}{2\pi mc} S_z. \quad (18.14)$$

We see that the magnetic moment of the system has a positive sign, i.e. that the system having an intrinsic angular momentum is paramagnetic. The ratio of the magnetic moment to the orbital momentum is equal to $eh/4\pi mc$. For the spin moment this ratio is twice as large.

§ 19. Paramagnetic susceptibility

Let us now consider the behaviour of a system containing a large number of atoms or molecules in an external magnetic field. The magnetic field has an orienting effect on the magnetic moments of the atoms, tending to set them along the field. The thermal motion disturbs such a regular arrangement of the moments. As a result of the competition of these processes a certain mean arrangement of the directions of the magnetic moments with respect to the direction of the field is established. To this mean arrangement of the elementary magnetic moments there corresponds the mean magnetic moment of the entire system.

Let us find the resulting magnetic moment of a system of atoms according to the general formula. We assume that any interaction between the magnetic moments is absent and each magnetic moment turns freely in the external field. In order that this assumption may be valid it is necessary that the mean distance between the atom is sufficiently large. Examples of such systems will be given below. If the initial assumption is fulfilled, then each particle can be considered as an individual subsystem having an energy in the

field given by formula (18.9). The mean value of the magnetic moment of the particle can be found by means of formula (18.1).

For the partition function we have

$$z = \sum \exp \left[-\frac{\epsilon_i^{(0)} - (eh/4\pi mc)L_z H}{kT} \right] \Omega(\epsilon_i). \quad (19.1)$$

The summation is to be carried out over all values of the energy of the sub-system. Since in the magnetic field $2L + 1$ adjacent energy levels arise in place of one energy level, the summation in (19.1) is carried out over all energy levels i and, within the limits of a given level, also over all sub-levels which are defined by formula (18.9) and which differ from each other by the discrete values of the magnetic energy $ehL_z H/4\pi mc$, since L_z takes on a discrete series of values. The expression (19.1) can essentially be simplified if it is taken into account that in atomic systems the spacing between levels is very large in comparison with the thermal energy kT . Because of this the terms of the sum will decrease rapidly and one can confine oneself to the first term which refers to the ground state energy ϵ_0 . In the magnetic field the latter level splits into $2L + 1$ or $2S + 1$ sub-levels, depending on which of the quantities, L or S , differs from zero in the ground state. In the first case one can write

$$z = \exp \left[-\frac{\epsilon_0}{kT} \right] \sum \exp \left[\frac{eh}{4\pi mc} \frac{L_z H}{kT} \right] \Omega(\epsilon_0), \quad (19.2)$$

where the sum is taken over the sub-levels mentioned.

In what follows we shall consider two limiting cases:

$$\frac{eh}{4\pi mc} L_z H \gg kT \quad (19.3)$$

and

$$\frac{eh}{4\pi mc} L_z H \ll kT. \quad (19.4)$$

Because of the smallness of the magnetic moment, the condition (19.4) is fulfilled for available fields at a not too low temperature. If condition (19.4) is fulfilled, then the exponent in (19.2) can be expanded in a series and one can restrict oneself to terms which are of low order in the field:

$$z = \exp \left[-\frac{\epsilon_0}{kT} \right] \Omega(\epsilon_0) \sum_{L_z = -L}^{L_z = L} \left(1 + \frac{eh}{4\pi mc} \frac{L_z H}{kT} + \frac{1}{2} \left(\frac{eh}{4\pi mc} \right)^2 \left(\frac{L_z H}{kT} \right)^2 + \dots \right).$$

Upon summing, we have

$$\sum 1 = 2L + 1,$$

$$\sum L_z = 0,$$

$$\sum L_z^2 = \frac{1}{3} L(L+1)(2L+1).$$

We then obtain

$$z = \exp \left[-\frac{\epsilon_0}{kT} \right] (2L+1) \Omega(\epsilon_0) \left[1 + \frac{1}{6} \left(\frac{eh}{4\pi mc} \right)^2 \frac{L(L+1)H^2}{(kT)^2} \right]. \quad (19.5)$$

The partition function of an entire system consisting of mutually independent particles is obviously equal to $Z = z^N$.

According to (18.1) the mean magnetic moment of the entire system is

$$M = kT \frac{\partial \ln Z}{\partial H} = NkT \frac{\partial \ln z}{\partial H} \approx \frac{1}{3} \left(\frac{eh}{4\pi mc} \right)^2 \frac{L(L+1)NH}{kT}. \quad (19.6)$$

The paramagnetic susceptibility referred to N molecules has the form

$$\chi = \frac{M}{H} = \frac{1}{3} \left(\frac{eh}{4\pi mc} \right)^2 \frac{L(L+1)N}{kT}. \quad (19.7)$$

If the magnetic moment is due not to the orbital momentum but to the spin, the instead of (19.6) we have by virtue of (18.14)

$$M = \frac{1}{3} \left(\frac{eh}{2\pi mc} \right)^2 \frac{S(S+1)NH}{kT}, \quad (19.8)$$

while instead of (19.7) we obtain

$$\chi = \frac{1}{3} \left(\frac{eh}{2\pi mc} \right)^2 \frac{S(S+1)N}{kT} \quad (19.9)$$

If, finally, the system has both orbital momentum and spin, then one obtains for χ an analogous expression but with the coefficient in the numerator having a value which is intermediate between (19.7) and (19.9).

In all the cases considered the paramagnetic susceptibility turns out to be inversely proportional to the temperature T . Besides the temperature, the magnetic susceptibilities (19.7) and (19.9) contain only a constant factor made up of universal constants and the value of the orbital momentum or spin respectively.

Passing on to the second limiting case where the condition (19.3) is satisfied, we see that of all the terms of the sum in (19.1) it is necessary to retain only one term corresponding to the value of the magnetic part of the energy $ehLH/4\pi mc$. Other terms of the sum which contain the components $eh(L-1)H/4\pi mc$, $eh(L-2)H/4\pi mc$ and so on will be much smaller than the first. The reason for this is clear: the inequality (19.3) means that the energy of orientation in the magnetic field is large in comparison with the thermal energy and that all magnetic moments will be oriented along the field, so that their projection L_z will be equal to L , i.e. a total saturation will ensue.

Dropping all terms except the first in the sum (19.2), we find

$$z = \exp \left[-\frac{\epsilon_0}{kT} \right] \exp \left[\frac{eh}{4\pi mc} \frac{LH}{kT} \right], \quad (19.10)$$

whence

$$M = NkT \frac{\partial \ln z}{\partial H} = \frac{ehN}{4\pi mc} LH, \quad (19.11)$$

or

$$M = \frac{ehN}{2\pi mc} SH \quad (19.12)$$

(depending on the nature of the magnetic moment). Thus total saturation ensues, and all moments are set up along the field. In the intermediate case one can obtain a general formula for the dependence of the magnetic moment on the field.

Considering now comparison with experiment, we note that all quantities

in formulae (19.7) and (19.9) are known. Hence calculated susceptibilities can be compared directly with experimental values for systems for which the initial assumption (the absence of interaction between particles possessing a magnetic moment) is fulfilled. It should be noted that the number of such systems is rather small. Most atoms and molecules in the ground state have an orbital momentum and spin equal to zero ($L=S=0$). Substances which in the ground state possess a magnetic moment and a paramagnetic susceptibility also have a diamagnetic susceptibility. However, the latter constitutes a relatively small (although sometimes quite appreciable) part of the paramagnetic susceptibility. To obtain the true value of the paramagnetic susceptibility it is necessary to add to its measured value the value of the diamagnetic susceptibility. It is not a simple problem to obtain atomic paramagnetic substances in the gaseous state and measure their susceptibilities. Nevertheless, the susceptibility of the vapour of K has been measured, and these measurements led to a value of the paramagnetic susceptibility of $0.38T^{-1}$, which is in agreement with the theoretical value of $0.37T^{-1}$. The accuracy of the measurements is not high, and they cannot be used for a complete verification of formula (19.9).

The most suitable objects for the verification of the theory of paramagnetic susceptibility described above are:

(1) Aqueous solutions or the solid crystalline hydrates of salts containing ions with an orbital angular momentum or spin differing from zero. Such ions are those of the elements of the rare-earth group and the transition elements of the iron group in solutions or crystalline hydrates. In aqueous solutions and crystalline hydrates paramagnetic ions are separated from each other by a large number of water molecules. Hence the energy of interaction between them is very small. The agreement of theory with experiment turns out to be excellent. This is confirmed in the example of gadolinium Gd^{3+} . For this ion $L=0$ and $S=\frac{7}{2}$. Its magnetic moment and susceptibility are expressed by formulae (19.8) and (19.9). The theoretical dependence of the magnetic moment on the quantity $\mu H/kT$ is shown in fig. IV.8 by a solid line. The measured values of M are shown by dots.

(2) The molecules of paramagnetic gases (O_2 , NO etc.). The electric fields of molecules have not spherical symmetry; their angular momentum has no fixed value and the mean value is equal to zero. If, however, a molecule possesses a spin different from zero, then it has a magnetic moment which is connected with the spin by the relation (18.14). The number of molecules having a spin different from zero is relatively small. As an example one can quote the oxygen molecule O_2 whose spin is $S=1$. For the magnetic susceptibility of 1 cm^3 of oxygen in normal conditions we obtain from formula

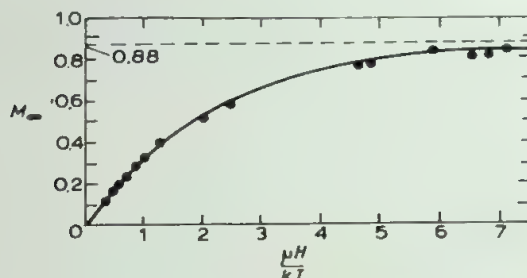


Fig. IV.8

(19.9) the value $\chi = 0.142 \times 10^{-6}$. The measured value is equal to $\chi = 0.143 \times 10^{-6}$. This agreement is excellent.

We cannot go into a discussion here of more complex cases of molecules having a non-zero projection of the angular momentum onto the symmetry axis, and cases where one cannot consider the spin of the molecule to be oriented freely in space *.

§20. Ferromagnetism. Spontaneous magnetization and hysteresis

Ferromagnetism represents a specific phenomenon which occurs only in the solid phase and for a relatively limited range of substances.

As we have stressed earlier, the magnetic properties of ferromagnetics differ essentially from those of other bodies. For ferromagnetics not only is there no proportionality between the vectors \mathbf{B} and \mathbf{H} but the induction is a complex and ambiguous function of the field,

$$\mathbf{B} = f(\mathbf{H}) .$$

The value of $f(\mathbf{H})$ depends on the history of magnetization.

As the external field \mathbf{H} increases the induction of the ferromagnetic sample increases according to a curve whose characteristic shape is shown

* See, for example, E.Bloch, *Molekularnaya teoriya magnetizma* (Molecular theory of magnetism) (GTTI, 1936); S.V.Vonsovskii, *Sovremenoe uchenie o magnetizme* (Contemporary teaching on magnetism) (Gostekhizdat, Moscow, 1952); D.H.Martin, *Magnetism in solids* (Iliffe, London, 1967).

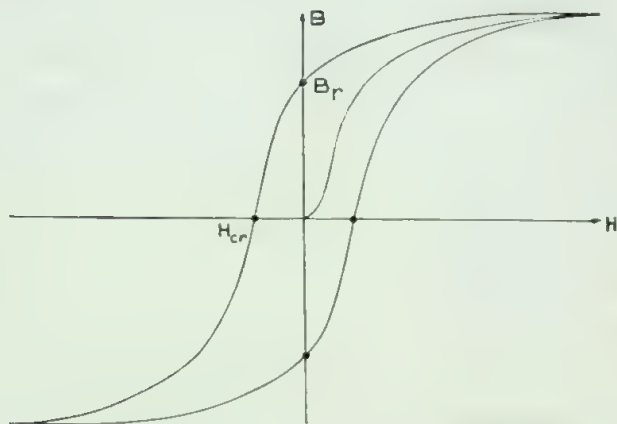


Fig. IV.9

in fig. IV.9. When the field is switched off the induction decreases but does not reach zero, so that a residual magnetization remains in the body, which exists even in the absence of the external field. This magnetization can be reduced to zero by changing the direction of the external field. The repetition of the process proceeds according to a characteristic cycle which is also shown in fig. IV.9 and which is called the hysteresis cycle.

The existence of such a connection between \mathbf{B} and \mathbf{H} means that in the relation

$$\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M},$$

the mean magnetic moment (magnetization) of the body must be considered as a quantity determined by the state of the body as a whole. This means that the relation between \mathbf{M} and \mathbf{H} for a ferromagnetic body has a complex character and that the definition of \mathbf{H} does not in itself determine the magnitude of the magnetization. We shall discuss below the properties of magnetization for ferromagnetics and its dependence on the field \mathbf{H} and on the temperature. However, it should first of all be stressed that, in contrast to diamagnetic and paramagnetic substances, the magnetization in ferromagnetic substances is not determined by the external field but is an internal parameter of the system.

A spontaneous magnetization can exist in a ferromagnetic body in a state

of statistical equilibrium. The state of the body will be completely characterized by the definition of its internal parameters, for example, the temperature, pressure and spontaneous magnetization. In the equilibrium state the thermodynamic potential $G(p, T, \mathbf{M})$ must be a minimum.

In the presence of an external magnetic field the magnetization of the body will be a function of the applied field. The form of this function, for given properties of the body, is determined by equilibrium conditions in such a way that

$$G(p, T, \mathbf{M}(\mathbf{H})) \rightarrow \min. \quad (20.1)$$

The numerical value of the magnetization in ferromagnetic bodies is very large. If the magnetic susceptibility is formally defined as

$$\kappa = (\partial \mathbf{M} / \partial \mathbf{H})_{\mathbf{H} \rightarrow 0}, \quad (20.2)$$

then it turns out that its values reach orders of magnitude of $10^5 - 10^6$.

As we have already stressed, the phenomenon of ferromagnetism can occur only in solid bodies. It turns out that the magnetization in single crystals possesses an important property of anisotropy. The magnetization has different values in different directions in the crystal. In the so-called directions of easy magnetization it has a large value, while in other directions it has a smaller value for a given value of the strength of the magnetizing field.

The ferromagnetic properties of matter depend strongly on the temperature. As the temperature increases the spontaneous magnetization decreases and at a certain temperature, which is characteristic of a given substance, it reduces to zero. This temperature is called the magnetic transition temperature or the Curie point θ . At the Curie point the substance loses its ferromagnetic properties. At temperatures lying above the corresponding temperature θ all ferromagnetics become paramagnetics. At $T > \theta$ the paramagnetic susceptibility depends on the temperature according to the law

$$\kappa \sim \frac{1}{T - \theta},$$

which is called the Curie-Weiss law.

Finally, the most important property of ferromagnetic bodies is the following: even if it is homogeneous from the macroscopic point of view, a ferromagnetic body is usually divided into regions of spontaneous magnetization or domains. The size of the domains is very large in comparison with

the size of molecules and each domain consists of a very large number of particles. Within the limits of each domain there is a magnetization different from zero. This means that the magnetic moments of all atoms within the domain are spontaneously oriented preferentially in one direction, forming a macroscopic region of spontaneous magnetization.

If the body were not exposed to the action of a field, then the magnetization of individual domains in the entire body would, on the average, be balanced, and the vector \mathbf{M} for the body as a whole would be equal to zero.

The action of the external field amounts to a re-orientation of the magnetization vectors of individual domains such that a magnetization of the entire body different from zero is produced. When the field is removed the moments of a part of the domains preserve the preferential orientation, "remember" the action of the field, and additional work by a field is needed to reform a system of domains with a random orientation of magnetic moments.

It should be emphasized that the existence of domains is not a hypothesis but represents a phenomenon which has been well investigated experimentally in detail.

All the properties of ferromagnetic bodies mentioned above fit well into the framework of the thermodynamic theory of ferromagnetism described below. However, the basic problem, the essence of the phenomenon of spontaneous magnetization, cannot of course find any explanation within the framework of the macroscopic theory.

The microscopic quantum theory of the phenomenon of spontaneous magnetization is described in Part VI of this book.

Passing on to the discussion of the thermodynamic theory * we have, first of all, to write an explicit expression for the thermodynamic potential of the system — the ferromagnetic single crystal in a state of statistical equilibrium in an external field \mathbf{H} .

We shall consider the behaviour of ferromagnetism near the Curie point, when the magnetization is relatively small.

Let $G_0(p, T, \mathbf{M})$ be the thermodynamic potential of the body in the presence of an external field. The basic equality for the thermodynamic potential in a field reads

$$dG = dG_0 - \mathbf{H} \cdot d\mathbf{M}. \quad (20.3)$$

* In the subsequent exposition in this section we follow L.D. Landau and E.M. Lifshitz, *Electrodynamics of continuous media* (Pergamon Press, Oxford, 1960). Our potential φ corresponds to the quantity $\varphi + (H^2/8\pi)$ in the book quoted.

The difference between this relation and the analogous expression (18.2) lies in the fact that the independent variable is the mean magnetic moment of the ferromagnetic body, \mathbf{M} . Hence, integrating (20.3) with respect to \mathbf{M} for a given external field, one can write

$$G(p, T, \mathbf{M}) = G_0(p, T, \mathbf{M}) - \mathbf{M} \cdot \mathbf{H}. \quad (20.4)$$

The thermodynamic potential G_0 depends on the magnitude as well as on the orientation of the vector \mathbf{M} with respect to the axis of the single crystal.

The effect of anisotropy is relatively small. Hence $G_0(p, T, \mathbf{M})$ can be written in the form

$$G_0(p, T, \mathbf{M}) = G_1(p, T, M) + G_2(p, T, M), \quad (20.5)$$

where the second term takes into account the effect of anisotropy. The first, basic, term depends only on the absolute value of the magnetization vector.

We shall confine ourselves to the simplest case where the crystal has only one symmetry axis which is the axis of easy magnetization. We choose for this the z -axis. For the magnetization along the z -axis, i.e. $M_z = M$, the thermodynamic potential G_0 must have its smallest value. In this case the dependence of G_2 on M can be written in the form

$$G_2(p, T, \mathbf{M}) = \text{const} \cdot (M_x^2 + M_y^2) = \text{const} \cdot (M^2 - M_z^2) = \beta M_z^2 = \beta M^2 \sin^2 \theta. \quad (20.6)$$

where βM_z^2 which does not depend on angle can always be included in G_1 . In the small quantity G_2 we have written out only the first terms of the expansion in powers of M . Here β is a positive constant, and θ is the angle between the z -axis and the vector \mathbf{M} . The linear term in this expansion must drop out. Indeed, the magnetic moment, which is proportional to the velocity of the particles, changes sign under the substitution $t \rightarrow -t$. The thermodynamic potential of a system in an equilibrium state is obviously invariant under this substitution. Hence the expansion must contain a combination of the quadratic terms M_x^2 , M_y^2 , M_z^2 . This combination must be chosen in such a way that G_0 should have a minimum for the magnetization along the z -axis. Since the z -axis is the symmetry axis, the components M_x^2 and M_y^2 must figure symmetrically in the expansion. The expression (20.6) satisfies all these requirements. Since $\beta > 0$, the z -axis is indeed the axis of easy magnetization. The potential G_0 has its minimum value ($G_0 = 0$) if the vector \mathbf{M} is directed along the z -axis ($\theta = 0$).

The total thermodynamic potential in the external field has the form

$$G = G_1 - \mathbf{M} \cdot \mathbf{H} + \beta M^2 \sin^2 \theta. \quad (20.7)$$

This formula shows that the magnetization \mathbf{M} will be oriented by the external field on the one hand, and by the natural direction of easy magnetization on the other hand. As a result, the vector \mathbf{M} will set at a certain angle θ_{\min} with respect to the z -axis for which the thermodynamic potential has a minimum. If the plane in which \mathbf{H} and the axis of easy magnetization lie is chosen as the (xz) -plane, then the vector \mathbf{M} will obviously lie in this plane.

We write (20.7) in components

$$G = G_0 - MH_x \sin \theta - MH_z \cos \theta + \beta M^2 \sin^2 \theta.$$

We shall find the equilibrium orientation, determined by the condition of the minimum of G , from the condition

$$\partial G / \partial \theta = 2\beta M^2 \sin \theta \cos \theta - MH_x \cos \theta + MH_z \sin \theta = 0.$$

Hence we easily obtain

$$(2\beta M \sin \theta - H_x)^2 (1 - \sin^2 \theta) = H_z^2 \sin^2 \theta. \quad (20.8)$$

This is an equation of the fourth power with respect to the quantity $\sin \theta$. It has either two or four real roots, depending on the values of H_x and H_z (for given M and β).

In the first case one value of the root corresponds to the angle θ_{\min} for which G has its minimum value. This is the equilibrium orientation of \mathbf{M} . The second value of the root leads to the maximum value of G , i.e. to a thermodynamically unstable orientation of the magnetization.

In the second case there are two minimum and two maximum values of the angle θ .

One of these minima θ_{\min} leads to the smallest value of the thermodynamic potential. The corresponding orientation of the vector \mathbf{M} is the equilibrium orientation. The second minimum corresponds to a metastable state of the crystal. For this orientation the thermodynamic potential is smaller than for any neighbouring orientations but larger than that for the orientation at the angle θ_{\min} .

The existence of metastable states allows one to understand qualitatively the origin of residual magnetization and hysteresis.

The equilibrium state corresponds to a total magnetization of the crystal equal to zero in the absence of an external field. On the contrary, in the metastable state the magnetic moment of the body as a whole is different from zero in the absence of a field. If in the process of magnetization the crystal is brought into the metastable state with a certain total magnetization $\int \mathbf{M}(\mathbf{H})dV$ and thereupon the external field is removed, then the system will remain for a very long time (practically indefinitely) in the state with the total magnetization $\int \mathbf{M}(\mathbf{H} \rightarrow 0)dV$. From formula (20.8) there also follows, in principle, the possibility of the existence of domains. Assume that the external field is oriented perpendicularly to the axis of easy magnetization, i.e. $H_z = 0$. Then (20.8) goes over into the equation

$$2\beta M \sin \theta - H = 0. \quad (20.9)$$

If $H < 2\beta M$, then (20.9) has two solutions

$$\theta_1 = \arcsin \frac{H}{2\beta M}, \quad \theta_2 = \arcsin \left(\frac{H}{2\beta M} - \pi \right).$$

To these solutions there corresponds one and the same minimum value of the thermodynamic potential, so that both of these are equilibrium solutions. However, the values of the magnetization \mathbf{M} will in these cases be respectively

$$M_x^{(1)} = M \sin \theta_1, \quad M_x^{(2)} = M \sin \theta_2 = -M \sin \theta_1 = -M_x^{(1)}. \quad (20.10)$$

Thus, in the equilibrium state two opposite orientations of the magnetization vector are possible.

If the entire crystal is divided into alternating regions with magnetization $M_x^{(1)}$ and $M_x^{(2)}$, then its thermodynamic potential will be a minimum and its state will be an equilibrium state.

We see that from the formal thermodynamic theory there follows the possibility of the existence of residual magnetization (and hysteresis) and the domain structure of the crystal.

The actual realization, magnitude and form of the domains depend on many factors.

For details we refer the reader to the book of L.D. Landau and E.M. Lifshitz and to the specialized literature of this field.

It goes without saying that the thermodynamic theory does not give and cannot give any answer to the basic question of the nature of spontaneous magnetization. The existence of the anisotropy energy is also postulated in this theory.

Let us now consider the dependence of the magnetic properties of ferromagnetics on the temperature near the Curie point. At the Curie point the ferromagnetic and paramagnetic states of the substance are in statistical equilibrium. This equilibrium is a phase equilibrium, and each of the states represents a phase, respectively the ferromagnetic phase and the paramagnetic phase.

The thermodynamic potential of the ferromagnetic phase is given by formula (20.5). Near the Curie point one can neglect the small term G_2 and expand G_1 in a series in powers of a small quantity (the magnetization \mathbf{M}) which reduces to zero at the point $T = \theta$. Then we have

$$G_{\text{ferro}}(p, T, \mathbf{M}) = G_0(p, T) + aM^2 + bM^4 - \mathbf{M} \cdot \mathbf{H}, \quad (20.11)$$

to terms of the fourth order of a small quantity.

The odd powers of \mathbf{M} drop out of the expansion: under the substitution ($t \rightarrow -t$) \mathbf{M} changes sign whereas G does not.

At the Curie point $T = \theta$ the coefficient a reduces to zero. Near the Curie point one can write

$$a \sim (T - \theta), \quad (20.12)$$

so that $a > 0$ above the Curie point and $a < 0$ below the Curie point. Assume that $b > 0$. As will be seen from what follows, for such a choice of the signs of a and b the thermodynamic theory well describes the ferromagnetic-paramagnetic phase transition.

The equilibrium condition reads

$$\partial G / \partial M = 2aM^2 + 4bM^3 - H = 0. \quad (20.13)$$

If there is no magnetic field, then in the equilibrium state either

$$M = 0, \quad (20.14)$$

or

$$a + 2bM^2 = 0. \quad (20.15)$$

Above the Curie point $a > 0$ and the condition of the minimum of the thermodynamic potential is the equality (20.14). Below the Curie point the equality (20.15) can be satisfied (since here $a < 0$) and

$$M = \left(\frac{|a|}{2b} \right)^{\frac{1}{2}} \sim (\theta - T)^{\frac{1}{2}}.$$

At the Curie point M reduces to zero and the thermodynamic potentials of the ferromagnetic and paramagnetic states are equal to each other. As is easily seen, the entropy does not change in the phase transition. Indeed, above the Curie point

$$S_{\text{para}} = - \left(\frac{\partial G_0}{\partial T} \right)_p.$$

Below the Curie point, by virtue of the relations (20.12) and (20.15), we have

$$\begin{aligned} S_{\text{ferro}} &= - \left(\frac{\partial G_{\text{ferro}}}{\partial T} \right)_p = - \left(\frac{\partial G_0}{\partial T} \right)_p - M^2 \frac{\partial a}{\partial T} = - \left(\frac{\partial G_0}{\partial T} \right)_p - \text{const} \cdot M^2 = \\ &= - \left(\frac{\partial G_0}{\partial T} \right)_p - \text{const} \cdot (\theta - T) = S_{\text{para}} - \text{const} \cdot (\theta - T). \end{aligned}$$

Hence it is seen that at the Curie point, for $T = \theta$, the entropy remains continuous. However, the heat capacity of the system undergoes a jump. Indeed,

$$(C_p)_{\text{ferro}} = T \left(\frac{\partial S_{\text{ferro}}}{\partial T} \right)_p = \text{const} \cdot T + (C_p)_{\text{para}}.$$

The jump of the heat capacity for $T = \theta$ is

$$\Delta C_p = \text{const} \cdot \theta,$$

the heat capacity of the ferromagnetic phase being higher than that of the paramagnetic phase (for $b > 0$), which is in complete agreement with experiment. The phase transition considered is a typical example of a phase transition of the second kind (see §63 of Part III).

For $H \neq 0$ one can find the paramagnetic susceptibility from the condition of equilibrium. Namely, for $H \neq 0$ the condition (20.13) instead of (20.14) must be fulfilled. Differentiating it with respect to H , we find

$$\frac{\partial}{\partial H} \left(\frac{\partial G}{\partial M} \right) = \frac{\partial M}{\partial H} (2a + 612bM^2) - 1 = 0,$$

whence

$$\kappa = \frac{\partial M}{\partial H} = \frac{1}{2a + 612bM^2}.$$

For $T > \theta$, $M = 0$, the paramagnetic susceptibility, in accordance with the Curie-Weiss law, is equal to

$$\kappa = \frac{1}{2a} \sim \frac{1}{(T - \theta)}. \quad (20.16)$$

In conclusion we shall dwell briefly on the calculation of electromagnetic fields in the presence of ferromagnetics. The calculation of such fields is particularly important in technology, where ferromagnetic materials are widely used.

At first sight it may seem that the non-linear relation between \mathbf{B} and \mathbf{H} makes the calculation of the magnetic field very difficult. In practice, however, this is not so. As a matter of fact, the calculation of the field distribution outside the ferromagnetic requires the definition of the boundary condition at its surface. The use of accurate boundary conditions would associate the external problem with the field distribution inside the ferromagnetic, i.e. would make the problem very complex. However, if the ferromagnetic is formally considered as a medium with a magnetic susceptibility (20.2) and a magnetic permeability $\mu = 1 + 4\pi\kappa$ then one obtains for μ values which are higher by 5–6 orders than those for the external medium. Hence one can assume with a high degree of accuracy that $\mu \rightarrow \infty$, and the boundary condition (5.5) at the surface of the ferromagnetic assumes the simple form

$$\mathbf{H} \times \mathbf{n} = 0,$$

i.e. the lines of the magnetic field are normal to the surface. A ferromagnetic in a constant magnetic field turns out to be similar to a conductor in a constant electrostatic field. The calculation of the magnetic field inside ferromagnetic bodies turns out, as a rule, to be a very difficult problem.

§21. Superconductivity

As far back as 1911 Kamerlingh-Onnes established that the temperature dependence of the resistance of mercury differs fundamentally from that of normal metals.

As the temperature decreases the resistance ceases to depend on the temperature as in the case of normal metals, and its value is determined by the impurities which are present in the sample. However, when the temperature is reduced to $T = 4.1$ K the resistance of the metal falls abruptly (fig. IV.10) to zero. This phenomenon, the discontinuous vanishing of the resistance, is called the transition of the metal into the superconducting state, or briefly, the appearance of superconductivity. The temperature of this transition is called the superconducting transition temperature.

It is now established that superconductivity is a relatively widespread phenomenon. It is known that there are 23 metals which can go into the superconducting state. Superconductivity is observed also for a large number of alloys.

There is no doubt that the reduction of the resistance to zero corresponds to the transition of the metal into a new state. The resistance of all metals in the superconducting state is no more than 10^{-10} of the resistance immediately before the transition. Current in a ring made of a superconducting material circulates for an indefinitely long time without any sign of weakening. Hence it should be recognized that the resistance of superconductors is not simply very small but is exactly equal to zero. Electrons in superconductors can move without any hindrance.

Initially it was assumed that the metal in the superconducting state is an ideal conductor, i.e. a body with an infinitely large conductivity. According to Ohm's law, for the ideal conductor (i.e. a body with $\sigma \rightarrow \infty$) a finite value

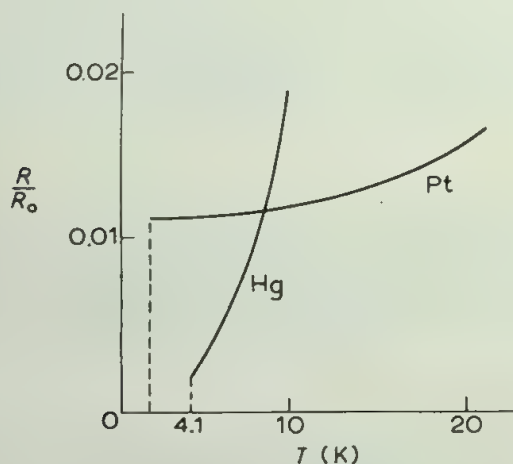


Fig. IV.10

of the current density \mathbf{j} corresponds to a field strength inside the conductor equal to zero:

$$\mathbf{E} = 0. \quad (21.1)$$

The absence of an electric field inside the superconductor implies definite magnetic properties. That is, from the Maxwell equation

$$\nabla \times \mathbf{E} = \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}$$

and from the fact that the field \mathbf{E} is equal to zero it follows that the magnetic induction in the superconductor has a constant value

$$\mathbf{B} = \text{const}. \quad (21.2)$$

The value of this constant is equal to the value of the magnetic induction in the superconductor at the moment of the transition into the superconducting state.

It turns out, however, that this inference is contrary to experiment. If a metal cylinder is placed in a magnetic field which is perpendicular to the axis of the cylinder, and if it is cooled below the transition temperature, then one can infer the character of the field inside the superconductor by its distribution near the sample. It turns out that the lines of the magnetic induction are pushed out of the superconductor (fig. IV.11). The induction inside the superconductor is equal to zero:

$$\mathbf{B} = 0. \quad (21.3)$$

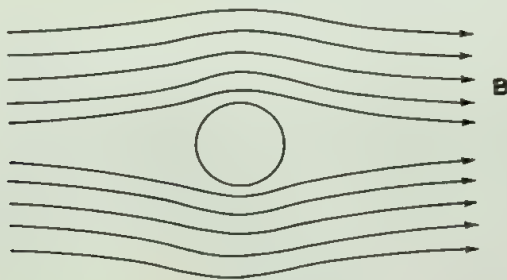


Fig. IV.11

From the boundary condition (5.5) the normal component of the external magnetic field $(H_e)_n$ (equal to the induction: $(H_e)_n = (B_e)_n$) is also equal to zero at the surface of the superconductor. In other words, the external magnetic field is tangential to a body which is in the superconducting state.

Eq. (1.16) shows that, if there is no magnetic induction in the body, then in the absence of an alternating electric field the total mean current in the volume of the body is zero:

$$\overline{\rho \mathbf{v}} = 0. \quad (21.4)$$

Thus inside the superconductor the total current density is zero. This means that in the surface layer of the superconductor a surface current circulates having a value such that the magnetic induction in the body is reduced to zero. In other words, the mean field of the surface currents compensates for the external magnetic field applied to the superconductor.

From the condition $\mathbf{B} = 0$ it follows that inside the superconductor the equations

$$\nabla \times \mathbf{E} = 0, \quad \nabla \cdot (\epsilon \mathbf{E}) = 0$$

hold, so that the electric field inside the superconductor is zero. The surface current in the superconductor is not associated with the action of the electric field. To determine the surface current density use can be made of the boundary condition (5.3). Namely, since there is no magnetic field in the superconductor, then (5.3) gives directly (for $\mathbf{H}^{(e)} = \mathbf{B}^{(e)}$)

$$\mathbf{j}_S = \frac{e}{4\pi} [\mathbf{n} \times \mathbf{B}^{(e)}],$$

where $\mathbf{B}^{(e)}$ is the magnetic induction vector outside the superconductor. If the body is simply connected, then the sum of the surface currents is equal to zero. On the other hand, in the case of a multiply connected body, for example a superconducting ring, the total current can be different from zero. The appearance of current in a superconductor may not be associated with the action of an impressed e.m.f. but can be produced by the action of a magnetic field. A surface current excited in a superconductor can circulate for an indefinitely long time without any weakening.

Experiment has shown that, if the magnetic field strength in the space which surrounds the superconductor exceeds a certain critical value H_{cr} , then the superconductivity in the sample vanishes. The destruction of the superconducting state and the appearance of a resistance take place in a discontinuous way.

The critical value of the magnetic field strength turns out to be a function of the temperature, which is given approximately by the empirical formula

$$H_{cr} = \text{const} \cdot (T_{cr} - T).$$

For $T = T_{cr}$, $H_{cr} = 0$, i.e. any field destroys the superconductivity at $T = T_{cr}$.

The consideration of the superconductor as a metal in a special superconducting state allows one to draw certain conclusions about the character of the transition into this state. Namely, the superconducting state should be considered as a particular phase of the substance, and the transition of the metal from the normal state into the superconducting state as a phase transition.

Let us consider the normal-state to superconducting-state phase transition occurring in an external field. In the superconducting state the mean total current $\overline{\rho \mathbf{v}}$ is equal to zero and it is impossible to separate from it the part which corresponds to the magnetic moment (as has been done in §3). Nevertheless, we introduce formally the magnetic field strength \mathbf{H}_i and the magnetic moment \mathbf{M} of the superconductor by the relation

$$\mathbf{B}_i = 0 = \mathbf{H}_i + 4\pi\mathbf{M}.$$

One can easily connect the field \mathbf{H}_i with the external magnetic field \mathbf{H}_e , if one considers a long cylindrical superconductor whose axis is directed along the field. Then, by virtue of the continuity of the tangential field component (which is in this case the same as the total field),

$$\mathbf{H}_i = \mathbf{H}_e$$

and

$$\mathbf{M} = -\mathbf{H}_e/4\pi. \quad (21.5)$$

We note that there corresponds to the equality (21.5) the value $\kappa = -1/4\pi$.

Thus, there corresponds to the superconducting state the magnetic permeability $\mu = 1 + 4\pi\kappa = 0$ and it is ideally diamagnetic.

According to (18.2) the thermodynamic potential of a body in the superconducting state can be written in the form

$$G_s(p, T, H) = G_s(p, T) - V \int_0^H \mathbf{M}(\mathbf{H}) \cdot d\mathbf{H}, \quad (21.6)$$

where $G_s(p, T)$ is the potential of the body in the superconducting state in the absence of an external magnetic field.

Substituting into (21.6) the value of $\mathbf{M}(\mathbf{H})$ according to formula (21.5), we obtain

$$G_s(p, T, H) = G_s(p, T) + V|\mathbf{H}_e|^2/4\pi. \quad (21.7)$$

When the external field reaches the critical value H_{cr} the superconducting state is destroyed. In this case the thermodynamic potential of the superconducting state increases so much, on account of the second term of (21.7), that the thermodynamic potential of the superconducting state turns out to be equal to the thermodynamic potential of the normal state $G_n(p, T)$. At the transition point the following relation is fulfilled:

$$G_s(p, T) + \frac{VH_{cr}^2}{8\pi} = G_n(p, T). \quad (21.8)$$

In the expression for the thermodynamic potential G_n we shall omit the additional term associated with the magnetic field, since it is very small for ordinary diamagnetic and paramagnetic metals.

Differentiating eq. (21.8) with respect to the temperature and making use of the definition of entropy (29.10) of Part III, we find

$$S_s - S_n = -\frac{\partial G_s}{\partial T} + \frac{\partial G_n}{\partial T} = \frac{VH_{cr}}{4\pi} \frac{dH_{cr}}{dT}, \quad (21.9)$$

where S_n and S_s are the entropies of the normal and superconducting states at the temperature T .

We see that, if $H_{cr} \neq 0$, i.e. if the transition occurs at $T < T_{cr}$, the entropy of the system changes in a discontinuous way. The phase transition is accompanied by the release of latent heat

$$Q = T\Delta S = \frac{TVH_{cr}}{4\pi} \frac{dH_{cr}}{dT}. \quad (21.10)$$

Experiment shows that in the transition from the normal state into the superconducting state heat is always released.

If the phase transition into the superconducting state occurs without field (i.e. if $H_{cr} = 0$), then

$$S_s = S_n ,$$

and the latent heat of transition is absent.

We can easily find the discontinuity of the heat capacity if we differentiate (21.9) with respect to the temperature. We then have

$$\left(\frac{\partial S_s}{\partial T}\right)_V - \left(\frac{\partial S_n}{\partial T}\right)_V = \frac{V}{4\pi} \left(\frac{\partial H_{cr}}{\partial T}\right)^2 + \frac{V}{4\pi} H_{cr} \left(\frac{\partial^2 H_{cr}}{\partial T^2}\right). \quad (21.11)$$

In the absence of an external field we find a simple expression for the jump of the heat capacity in the transition from the normal state into the superconducting state:

$$\Delta C_V = T \left[\left(\frac{\partial S_s}{\partial T}\right)_V - \left(\frac{\partial S_n}{\partial T}\right)_V \right] = \frac{VT}{4\pi} \left(\frac{\partial H_{cr}}{\partial T}\right)^2. \quad (21.12)$$

Thus, the phase transition considered represents a phase transition of the second kind. The derived macroscopic (thermodynamic) relations and a number of other relations which we shall not touch upon * are in good agreement with experimental data. However, although the thermodynamic theory of superconductivity successfully describes a number of properties of superconductors, it leaves open the basic problem: the problem of the physical nature of superconductivity.

Until recently all attempts to create a microscopic theory of superconductivity in which the superconductivity would be associated with the properties of the electron gas in the metal failed. They were only crowned with success in 1958. The contemporary theory of superconductivity will be expounded in Part V.

In the development of the theory of superconductivity an important role was played by experimental data which showed that there was a connection between the phenomenon of superconductivity and the character of interaction of the electrons of the superconductor with its crystal lattice. This was the so-called isotope effect, which was discovered in 1950 by Maxwell, Reynolds and others. It was discovered that the critical temperature T_{cr} depends on the isotope of a given element of which the lattice is made up, i.e. on the mass of the ions of the lattice. The relation between the mass M_{ion}

* See L.D. Landau and E.M. Lifshitz, *Electrodynamics of continuous media* (Pergamon Press, Oxford, 1960).

of the ions of the lattice and the critical temperature of the transition is given by the empirical formula

$$M_{\text{ion}}^{\frac{1}{2}} T_{\text{cr}} = \text{const} , \quad (21.13)$$

where the constant has a definite value for each element.

In conclusion we point to the analogy existing between the superconducting state of a metal and the superfluid state of liquid helium II. The superconductivity can be considered as the superfluidity of the electron gas which can move without hindrance in the crystal lattice of the metal (see Part V).

Quasistationary Electromagnetic Fields

§22. Conditions of quasistationarity

We now go on to the study of electromagnetic fields changing in time. It turns out that there is a wide frequency range for which Maxwell's equations allow an essential simplification.

In §19 of Part I we considered the case of a slow movement of charges where it was possible to completely disregard the retardation in the system and to assume the electromagnetic field to propagate with an infinitely large velocity.

In the results of §26 of Part I the condition for disregarding the retardation reduced to the requirement

$$T \gg \tau = L/c, \quad (22.1)$$

where T is the period of motion in the system, τ is the delay time, and L is the geometrical extent of the region in which the electromagnetic perturbations are considered.

In macroscopic electrodynamics there is a number of important problems in the treatment of which one can assume condition (22.1) to be fulfilled. In the approximation (22.1), the velocity of propagation of electromagnetic

disturbances can be assumed to be infinitely large within the limits of the system considered. In this case the values of the fields at a given point will be in phase with those at any other point inside the system. It is clear that the possibility of neglecting the retardation in the system essentially simplifies the study of the corresponding electromagnetic fields.

In Part I we have called electromagnetic fields in which the phenomenon of retardation can be disregarded quasistationary fields. In studying electromagnetic phenomena in matter, quasistationary fields, must, in addition to the requirement (22.1) which we shall call the first condition of quasistationarity, satisfy two more restrictions which we are going to derive.

The condition (22.1) obviously imposes a restriction upon the period T of the variation of the electromagnetic field (or upon the frequency $\omega = 2\pi/T$). That is, the periods T must be sufficiently large (and the frequencies sufficiently low) for a given size of system.

For relatively low frequencies of the electromagnetic fields in conductors the following condition is always fulfilled:

$$\frac{4\pi}{c} \mathbf{j} \gg \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t}. \quad (22.2)$$

As a matter of fact, it can be rewritten in the form

$$\sigma \mathbf{E} \gg \epsilon \frac{\partial \mathbf{E}}{\partial t} \approx \epsilon \omega \mathbf{E},$$

or

$$T \gg \epsilon / \sigma. \quad (22.3)$$

When the inequality (22.2), which should be called the second condition of quasistationarity, is satisfied in the region of space occupied by the conductors one can disregard the displacement current in comparison with the conduction current.

The third condition of quasistationarity is the requirement that the quantities characterizing the properties of matter — σ , ϵ and μ — should have the same value as in constant fields. We shall see in Part VI that the requirement about σ reduces to the need for the period T to be substantially longer than the mean time between collisions of electrons in the metal:

$$T \gg \lambda / \bar{v}, \quad (22.4)$$

where λ is the mean free path, and \bar{v} is the mean velocity of electrons in the metal. For smaller values of T the field manages to change its value appreciably between collisions, which obviously has an effect on the free path and, in the end, on the value of σ .

We shall consider later the dependence of ϵ on the frequency of the external field.

Numerical estimates show that conditions of quasistationarity for ordinary macroscopic systems containing metals as conductors are satisfied up to frequencies lying in the infrared part of the spectrum. The whole set of conditions determining quasistationary fields turns out to be fulfilled for a wide range of phenomena which have the common name "alternating currents". Alternating currents or low-frequency currents have a very wide application in technology and laboratory practice. This determines the practical importance of the theory of quasistationary processes.

The equations of the quasistationary electromagnetic field have the form

$$\left. \begin{aligned} \nabla \times \mathbf{H} &= \frac{4\pi}{c} \mathbf{j}, \\ \nabla \cdot \mathbf{B} &= 0, \end{aligned} \right\} \quad (22.5)$$

$$\left. \begin{aligned} \nabla \times \mathbf{E} &= -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \\ \nabla \cdot \mathbf{D} &= 4\pi\rho. \end{aligned} \right\} \quad (22.6)$$

In the equation for the curl of the magnetic field, on the basis of (22.2) we have dropped the term expressing the displacement current. The constitutive relations are:

$$\mathbf{B} = \mu \mathbf{H}; \quad \mathbf{D} = \epsilon \mathbf{E}; \quad \mathbf{j} = \sigma(\mathbf{E} + \mathbf{E}^{\text{imp}}). \quad (22.7)$$

The equation of continuity for quasistationary fields can be written in the form

$$\nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t} = \nabla \cdot \mathbf{j} + \frac{\partial}{\partial t} \left(\frac{\nabla \cdot \mathbf{D}}{4\pi} \right) = \nabla \cdot \left(\mathbf{j} + \frac{1}{4\pi} \frac{\partial \mathbf{D}}{\partial t} \right) \approx \nabla \cdot \mathbf{j} = 0. \quad (22.8)$$

Thus, the difference between the equations of the quasistationary field and those of the field of stationary currents reduces only to taking account

of the phenomena of electromagnetic induction. It will be shown below that the possibility of disregarding the retardation in the system allows one to obtain the equations of the electromagnetic field for the case of a system of linear conductors in the form of equations in total derivatives, and not partial derivatives, with constant coefficients.

For this purpose it is necessary to go over to Maxwell's equations in integral form.

§23. The law of induction in moving conductors and media

First of all it is necessary to find the flux of induction appearing in Maxwell's equations written in integral form. For this we have to generalize the notion of the flux of induction to the case of moving current loops.

Let a current loop be moving in a magnetic field. We assume the velocity of motion to be constant in space and small in comparison with the velocity of light c . Let us find the change in the flux of induction through a conducting loop.

We have, obviously,

$$\frac{d\Phi}{dt} = \frac{d}{dt} \int \mathbf{B} \cdot d\mathbf{S}.$$

By definition

$$\frac{d}{dt} \int \mathbf{B} \cdot d\mathbf{S} = \lim_{\Delta t \rightarrow 0} \left[(\Delta t)^{-1} \left(\int_{S_2} \mathbf{B}(t+\Delta t) \cdot d\mathbf{S} - \int_{S_1} \mathbf{B}(t) \cdot d\mathbf{S} \right) \right], \quad (23.1)$$

where $\mathbf{B}(t+\Delta t)$ is the induction vector taken at instant $t + \Delta t$, and S_2 is the surface into which the surface S_1 goes over at the instant $t + \Delta t$. The vectors of the normal to the two surfaces are assumed to be oriented in one direction.

We apply the Gauss-Ostrogradsky theorem to a closed volume (fig. IV.12) consisting of surfaces S_2 and S_1 and a lateral surface Σ which is formed when the loop is displaced from the position S_1 to the position S_2 . Since $\nabla \cdot \mathbf{B} = 0$, i.e. \mathbf{B} is a vector having no sources, one can write

$$\begin{aligned} \oint \mathbf{B}(t+\Delta t) \cdot d\mathbf{S} = \\ = \int_{S_2} \mathbf{B}(t+\Delta t) \cdot d\mathbf{S} + \int_{\Sigma} \mathbf{B}(t+\Delta t) \cdot d\mathbf{S} - \int_{S_1} \mathbf{B}(t+\Delta t) \cdot d\mathbf{S} = 0. \end{aligned} \quad (23.2)$$

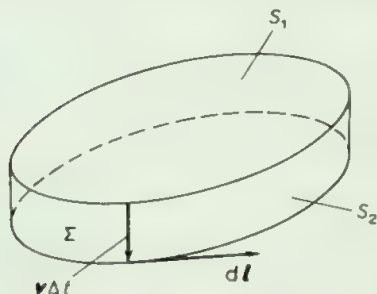


Fig. IV.12

Here the minus sign is associated with our choice of the orientation of the normal vector. For the lateral surface Σ one can, obviously, write (see fig. IV.12)

$$d\Sigma = (dl \times v) \Delta t,$$

where v is the velocity of motion of the current loop, and dl is an element of its length. Hence

$$\int_{\Sigma} \mathbf{B} \cdot d\Sigma = \int \mathbf{B} \cdot (dl \times v) \Delta t = \Delta t \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l}, \quad (23.3)$$

where the integral is taken with respect to the curve bounding the surface S_1 (i.e. with respect to the current loop). Further, to an accuracy of second order infinitesimal quantities, one can write

$$\int_{S_1} \mathbf{B}(t+\Delta t) \cdot d\mathbf{S} \approx \int_{S_1} \mathbf{B}(t) \cdot d\mathbf{S} + \Delta t \int_{S_1} \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S}. \quad (23.4)$$

From (23.3), (23.4) and (23.2) we find

$$\int_{S_2} \mathbf{B}(t+\Delta t) \cdot d\mathbf{S} \approx \int_{S_1} \mathbf{B}(t) \cdot d\mathbf{S} + \Delta t \int_{S_1} \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} - \Delta t \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l}.$$

Substituting this expression into (23.1) and passing over to the limit, we find

$$\frac{d\Phi}{dt} = \frac{d}{dt} \int \mathbf{B} \cdot d\mathbf{S} = \int_{S_1} \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} - \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l}. \quad (23.5)$$

Formula (23.5) shows that a change with time in the flux of induction through the current loop can occur either owing to a change in the induction vector, or owing to the motion of the current loop at a non-zero angle with respect to the lines of the field (so that the velocity vector is not parallel to \mathbf{B}). In other words, if the moving current loop intersects the lines of the induction vector \mathbf{B} , then a change in the induction flux Φ in time arises.

Passing, in the second integral, over to the integration with respect to the surface, we can write

$$\frac{d\Phi}{dt} = \int \left[\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) \right] \cdot d\mathbf{S}. \quad (23.6)$$

We now write Faraday's law of induction for the moving current loop in the form

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\frac{1}{c} \frac{d\Phi}{dt}, \quad (23.7)$$

where \mathbf{E} is the field strength in the conductor. It is clear that the total change in the flux of induction must remain on the right-hand side of (23.7) irrespective of the cause of this change.

The continuity of the tangential component of \mathbf{E} allows one to pass over from the conductor to a neighbouring contour lying in the medium outside the conductor. In this case one should not speak of the motion of the current loop but of the motion of the medium. The velocity \mathbf{v} denotes then the velocity of motion of a given point of the medium. One can pass over from (23.7) to the Maxwell differential equation in a moving medium.

Namely, writing by means of (23.5)

$$\oint \mathbf{E} \cdot d\mathbf{l} = \int (\nabla \times \mathbf{E}) \cdot d\mathbf{S} = -\frac{1}{c} \int \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} + \frac{1}{c} \int \nabla \times (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{S},$$

we have

$$\nabla \times \mathbf{E} = -\frac{1}{c} \left[\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) \right]. \quad (23.8)$$

We transform the expression in parentheses, writing on the basis of (I.45) and (I.18) that

$$\begin{aligned}\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) &= \frac{\partial \mathbf{B}}{\partial t} - (\mathbf{B} \cdot \nabla) \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{B} - \mathbf{v}(\nabla \cdot \mathbf{B}) + \mathbf{B}(\nabla \cdot \mathbf{v}) = \\ &= \frac{\partial \mathbf{B}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{B} = \frac{d\mathbf{B}}{dt}\end{aligned}$$

by virtue of (22.5) and the constancy of \mathbf{v} .

Then we find finally the value of $\nabla \times \mathbf{E}$ in the moving medium:

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{d\mathbf{B}}{dt}. \quad (23.9)$$

The importance of the results obtained and, in particular, Faraday's law of induction in the form (23.7), is associated with the fact that the motion of conductors in a magnetic field is in practice one of the basic methods of producing an e.m.f. (for example, in the dynamo).

§ 24. Maxwell's equations for quasistationary fields in integral form and their integration for the case of linear conductors

We can now consider the Maxwell system of equations for quasistationary fields in integral form by considering the general case of moving media:

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\frac{1}{c} \frac{d\Phi}{dt}, \quad (24.1)$$

$$\oint \mathbf{B} \cdot d\mathbf{S} = 0, \quad (24.2)$$

$$\oint \mathbf{H} \cdot d\mathbf{l} = \frac{4\pi I}{c}, \quad (24.3)$$

$$\oint \mathbf{D} \cdot d\mathbf{S} = 4\pi \int \rho \, dV. \quad (24.4)$$

According to (22.8), the continuity equation is written in the form

$$\oint \mathbf{j} \cdot d\mathbf{S} = 0. \quad (24.5)$$

Consider the case of linear conductors carrying current. For simplicity we confine ourselves first to one conducting loop. We assume that the impressed e.m.f. in the loop is defined; the value of the e.m.f. can depend on the time. We write the generalized Ohm's law in integral form as follows:

$$\oint \mathbf{j} \cdot \frac{d\mathbf{l}}{\sigma} = \oint \mathbf{E} \cdot d\mathbf{l} + \oint \mathbf{E}^{\text{imp}} \cdot d\mathbf{l}. \quad (24.6)$$

Since we are considering a linear contour and the continuity equation is valid, we can write it by means of (24.1), (15.4) and (15.6) in the form

$$IR = -\frac{1}{c} \frac{d\Phi}{dt} + \mathcal{E}(t). \quad (24.7)$$

Eq. (24.7), which connects the current in a loop with a given impressed e.m.f. $\mathcal{E}(t)$, is usually called Ohm's law for an alternating-current circuit.

If we now consider a system consisting of an arbitrary number N of loops carrying alternating current, then for each of these we can write

$$I_i R_i = -\frac{1}{c} \frac{d\Phi_i}{dt} + \mathcal{E}_i(t), \quad (24.8)$$

where Φ_i is the flux of induction through the i th conducting loop.

The magnetic field is determined from eqs. (24.2) and (24.3) which do not differ from the equations for direct current. Hence the relations obtained for direct current remain valid for all magnetic quantities. However, the equations for the magnetic field distribution and the current distribution in the circuit turn out to be interrelated: the flux of magnetic induction through the i th current loop is determined by the magnetic field distribution; the magnetic field distribution is determined by the currents in all the loops.

It is easily shown, however, that in quasistationary fields the fluxes of induction turn out to be connected with the current densities by linear relations with constant coefficients of the type

$$\Phi_i = c \sum_{k=1}^N L_{ik} I_k \quad (L_{ik} = L_{ki}). \quad (24.9)$$

For the proof we shall consider the simplest case of two current loops in a medium with magnetic permeability μ .

The current in the second loop (fig. IV.13) produces in the first loop a flux

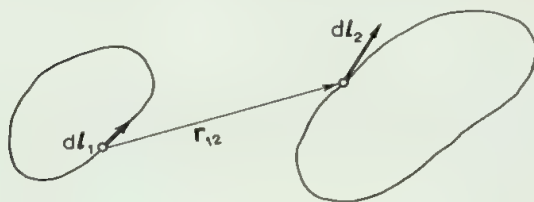


Fig. IV.13

$$\Phi_1(2) = \mu \int \mathbf{H} \cdot d\mathbf{S}_1 = \int (\nabla \times \mathbf{A}_2) \cdot d\mathbf{S}_1 = \oint \mathbf{A}_2 \cdot d\mathbf{l}_1, \quad (24.10)$$

where the index 1 refers to the first loop.

The vector potential of the magnetic field of the linear conductor carrying current I_2 is given by formula (17.4). Hence

$$\Phi_1(2) = \frac{\mu I_2}{c} \oint \oint \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r_{12}} = c L_{12} I_2, \quad (24.11)$$

i.e. the flux of induction in the first loop, which is produced by the current in the second loop, is proportional to the current I_2 in the latter. The factor of proportionality L_{12} turns out to be a constant which depends only on the properties of the medium and the relative position of the loops:

$$L_{12} = \frac{\mu}{c^2} \oint \oint \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r_{12}}, \quad (24.12)$$

where r_{12} is the distance between the elements of length $d\mathbf{l}_1$ and $d\mathbf{l}_2$. The quantity L_{12} is called the coefficient of mutual inductance.

Besides the flux of magnetic induction $\Phi_1(2)$, the flux produced by the current in the first loop itself

$$\Phi_1(1) = c L_{11} I_1, \quad (24.13)$$

passes through the first loop. The factor of proportionality L_{11} is called the coefficient of self-inductance. The meaning of the coefficient of self-inductance is the following. The flux of magnetic induction through a certain segment of the wire is obtained by summing the fluxes from its other segments. In all segments of the wire there flows the same current I_1 .

It is clear, however, that formula (24.12) cannot be used for the calculation of the coefficient of self-inductance. When the segment $d\mathbf{l}_2$ approaches the segment $d\mathbf{l}_1$ in the same conductor $r_{12} \rightarrow 0$, and the integral (24.12) diverges. The cause of this divergence lies in the fact that, if the segment $d\mathbf{l}_2$ is close to $d\mathbf{l}_1$, the transverse dimensions of the conductor become comparable with the distance r_{12} between these segments. This, in its turn, does not allow one to make use of the formula (17.4) which was found for the approximation of linear conductors. In other words, in considering the interrelation between nearby segments of one conductor the linear conductor approximation is undoubtedly incorrect. In later sections we shall see how one can find the coefficient of self-inductance without having recourse to the idea of the linear character of the conductor. In the meantime we can write the total flux of magnetic induction through the first conductor in the form

$$\Phi_1 = cL_{11}I_1 + cL_{12}I_2. \quad (24.14)$$

Analogously, the flux of magnetic induction through the second conductor is

$$\Phi_2 = cL_{22}I_2 + cL_{21}I_1.$$

It is easily shown that the equality $L_{21} = L_{12}$ holds. Indeed,

$$L_{21} = \frac{\mu}{c^2} \oint \oint \frac{d\mathbf{l}_2 \cdot d\mathbf{l}_1}{r_{21}} = \frac{\mu}{c^2} \oint \oint \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r_{12}} = L_{12}. \quad (24.15)$$

The equation (24.7) of the generalized Ohm's law for the first loop assumes the form

$$I_1 R_1 = -L_{11} \frac{dI_1}{dt} - L_{12} \frac{dI_2}{dt} + \mathcal{E}_1(t). \quad (24.16)$$

An analogous expression can be written for the second loop.

In the general case of an arbitrary number of loops the flux of magnetic induction through the loop i is expressed according to formula (24.9) in terms of currents in all the loops, the coefficient of self-inductance L_{ii} and the coefficients of mutual inductance L_{ik} . Substituting (24.9) into Ohm's law for alternating-current circuits, we find the whole set of equations which determine the current I_i in each of the N loops:

$$I_i R_i = - \sum_k L_{ik} \frac{dI_k}{dt} + \mathcal{E}_i(t) \quad (i=1,2,\dots,N). \quad (24.17)$$

If the impressed e.m.f. \mathcal{E}_i and the geometrical properties of the system of loops determining the whole set of coefficients L_{ik} are defined, then the integration of the system of linear differential equations (24.17) with constant coefficients allows one to determine the currents I_i in the loops. In this case the magnetic fields will be determined according to formulae (24.2) and (24.3).

Thus, corresponding to what was said at the end of §22, the equations of the quasistationary electromagnetic field for the case of a system of linear conductors reduce to equations in total derivatives with constant coefficients. The latter, however, as is known from the general theory of differential equations, reduce to a system of algebraic equations. We shall carry this out in following sections.

The densities of volume charges arising in certain cases in conductors are usually calculated by means of formula (24.4) from a given distribution of the electric field.

As an example of the applications of formulae (24.11) and (24.12) we shall calculate the coefficient of mutual inductance of two straight parallel linear conductors of length l . For definiteness we assume the currents in the conductors to be directed in one direction. We denote by h the distance between the conductors. Then, according to formula (24.12), we have

$$\begin{aligned} L_{12} &= \frac{\mu}{c^2} \int \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r_{12}} = \frac{\mu}{c^2} \int_0^l \int_0^l \frac{dx_1 dx_2}{[(x_1 - x_2)^2 + h^2]^{\frac{1}{2}}} = \\ &= \frac{\mu}{c^2} \int_0^l dx_1 \ln \left[\frac{(l - x_1) + [(l - x_1)^2 + h^2]^{\frac{1}{2}}}{-x_1 + [x_1^2 + h^2]^{\frac{1}{2}}} \right]. \end{aligned}$$

The calculation of the second integral is somewhat cumbersome. Integrating by parts, we obtain

$$L_{12} = \frac{\mu}{c^2} \left\{ l \ln \left[\frac{h}{-l + [l^2 + h^2]^{\frac{1}{2}}} \right] - 2([l^2 + h^2]^{\frac{1}{2}} - h) + l \ln \left[\frac{l}{-l + [l^2 + h^2]^{\frac{1}{2}}} \right] \right\}.$$

We can simplify this expression for two limiting cases: when $h \ll l$ or $h \gg l$.

In the first case we easily find, with an accuracy to the order of magnitude of h^2/l^2 ,

$$L_{12} \approx \frac{2\mu l}{c^2} \left[\ln \frac{2l}{h} - 1 + \frac{h}{l} \right].$$

In the other limiting case, with an accuracy to the order of magnitude of l^2/h^2 ,

$$L_{12} \approx \frac{\mu}{c^2} \frac{l^2}{h}.$$

In order to compare different methods of calculating L_{12} we shall find the coefficient of mutual inductance of two solenoids wound around a common toroidal core.

We assume the diameter D of the ring (proportional to the length of the solenoids) to be large in comparison with the diameter of the turns. The magnetic field inside the toroidal solenoid can be assumed to be uniform, and one can write

$$\oint \mathbf{H} \cdot d\mathbf{l} = 2\pi DH = \frac{4\pi}{c} n_1 I_1,$$

where n_1 is the number of turns in the first solenoid, and I_1 is the current in it. Whence

$$H = \frac{2n_1 I_1}{cD}. \quad (24.18)$$

The flux of induction through n_2 turns of the second solenoid (assuming that there is no leakage and that all lines of flux of the magnetic field produced by the first solenoid permeate the second solenoid) will be equal to

$$\Phi_{21} = \mu H n_2 S = \frac{2\mu n_1 n_2 S}{cD} I_1,$$

where S is the cross-section of the solenoid. From (24.11) we find

$$L_{21} = \frac{2\mu n_1 n_2 S}{c^2 D}.$$

§ 25. The energy of the magnetic field of a system of quasistationary currents

Let us find the energy of the magnetic field of a system of currents in the approximation of quasistationary fields.

As we have seen in § 7, the total energy (more precisely the free energy) of the magnetic field is

$$T_{\text{magn}} = \int \frac{\mathbf{B} \cdot \mathbf{H}}{8\pi} dV.$$

Here it has been assumed that there are no ferromagnetics in the volume V . The integration is carried out over all space.

Expressing \mathbf{B} in terms of the vector potential and making use of the first of eqs. (22.5), we have

$$\mathbf{B} \cdot \mathbf{H} = \mathbf{H} \cdot (\nabla \times \mathbf{A}) = \mathbf{A} \cdot (\nabla \times \mathbf{H}) + \nabla \cdot (\mathbf{A} \times \mathbf{H}) = \frac{4\pi}{c} (\mathbf{A} \cdot \mathbf{j}) + \nabla \cdot (\mathbf{A} \times \mathbf{H}).$$

Hence

$$T_{\text{magn}} = \frac{1}{2c} \int \mathbf{A} \cdot \mathbf{j} dV + \int \frac{\nabla \cdot (\mathbf{A} \times \mathbf{H})}{8\pi} dV. \quad (25.1)$$

In integrating over all space the last integral reduces to zero, and we find finally

$$T_{\text{magn}} = \frac{1}{2c} \int \mathbf{A} \cdot \mathbf{j} dV. \quad (25.2)$$

In formula (25.2) \mathbf{A} represents the vector potential produced at a given point by all the currents, and \mathbf{j} is the current density at the same point.

The integration over the entire volume can now be replaced by the integration over the volumes occupied by the current carrying conductors. Outside these volumes $\mathbf{j} = 0$ and the integral (25.2) vanishes. Hence

$$T_{\text{magn}} = \frac{1}{2c} \int \mathbf{A} \cdot \mathbf{j} dv, \quad (25.3)$$

where dv is a volume element of the conductor.

Let there be N conductors in the entire system. We divide the integral (25.3) into two parts, writing

$$T_{\text{magn}} = \frac{1}{2c} \sum_i \int \mathbf{A}_i \cdot \mathbf{j}_i dv_i + \frac{1}{c} \sum_{k>i} \int \mathbf{A}_k \cdot \mathbf{j}_i dv_i. \quad (25.4)$$

In the integral contained in the first sum, \mathbf{A}_i denotes the vector potential of the magnetic field produced in the volume element dv_i of a given conductor by the currents which are present in other elements of the same conductor.

In the integral in the second sum, \mathbf{A}_k represents the vector potential of the field produced in the volume element dv_i by the currents in all the other conductors. Consequently, the first sum represents the sum of the self energies of all the N conductors. The second sum gives the mutual energy of these conductors.

An actual calculation of the energy T_{magn} encounters great difficulties in the case of conductors which are placed in a medium which can be magnetized. Since the magnetic permeability of the conductors, generally speaking, differs from the permeability of the medium, it is necessary to know the distribution of the vector potential in a non-homogeneous medium in order to calculate the energy of the currents according to formula (25.4). Finding this distribution presents great difficulties. We shall therefore confine ourselves to the case where $\mu \approx \mu_0 \approx 1$ in the conductors as well as in the medium which surrounds them. In other words, we shall consider the conductors together with the medium which surrounds them as a quasi-homogeneous system. In this case the vector potential at any point of space is given by formula (17.4).

Substituting (17.4) into (25.4), we find

$$T_{\text{magn}} = \frac{\mu}{2c^2} \sum_i \iint \frac{\mathbf{j}'_i \cdot \mathbf{j}_i dv_i dv'_i}{r} + \frac{\mu}{c^2} \sum_{k>i} \iint \frac{\mathbf{j}_k \cdot \mathbf{j}_i dv_i dv_k}{r}, \quad (25.5)$$

where \mathbf{j}' is the density of the current in the volume element dv'_i producing the field in the volume dv_i , and r is the distance between dv_i and dv'_i .

Just as the energy of the electrostatic field in the case of a system of conductors reduces to the sum of the energies of the latter, the energy of the magnetic field given by formula (25.5) can be considered as the energy of the quasistationary currents.

We define the coefficients of self-inductance L_{ii} and the coefficients of mutual inductance L_{ik} in such a way that the following equality holds:

$$\begin{aligned}
 T_{\text{magn}} &= \frac{\mu}{2c^2} \sum_i \iint \frac{\mathbf{j}_i \cdot \mathbf{j}'_i dv_i dv'_i}{r} + \frac{\mu}{c^2} \sum_{k>i} \iint \frac{\mathbf{j}_k \cdot \mathbf{j}_i dv_i dv_k}{r} \equiv \\
 &\equiv \sum_i L_{ii} \frac{I_i^2}{2} + 2 \sum_{k>i} L_{ik} \frac{I_i I_k}{2}.
 \end{aligned} \tag{25.6}$$

The first sum represents the sum of the self energies of the currents in all the conductors, while the second sum represents the sum of the mutual energies.

Formula (25.6) is equally valid for linear and non-linear conductors.

In the case of linear conductors the definition of the coefficients of mutual inductance is the same as (24.12). Indeed, in this case

$$\frac{\mu}{c^2} \sum_{k>i} \iint \frac{\mathbf{j}_k \cdot \mathbf{j}_i dv_i dv_k}{r} = \frac{\mu}{c^2} \sum_{k>i} I_i I_k \iint \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r} = \sum_{k>i} I_i I_k L_{ik},$$

where the definition of L_{ik} is the same as (24.12).

The new definition of the coefficients of self-inductance is of fundamental importance for non-linear conductors. It is free of the difficulties associated with an indefinite increase of the integral as $r \rightarrow 0$ (see § 24). As $r \rightarrow 0$ the volume elements $dv dv'$ decrease more rapidly than r^{-1} , and the integral tends to zero. For this reason, as we shall see below in a practical example, formula (25.6) serves as the definition of the coefficients of self-inductance.

The energy T_{magn} of a set of linear conductors can be written in another form. Namely, in the case of linear currents the first term of (25.4) can be written in the form

$$\begin{aligned}
 T_{\text{magn}} &= \frac{1}{2c} \sum_i \int \mathbf{A}_i \cdot \mathbf{j}_i d\mathbf{S}_i \cdot d\mathbf{l}_i = \frac{1}{2c} \sum_i I_i \oint \mathbf{A}_i \cdot d\mathbf{l}_i = \\
 &= \frac{1}{2c} \sum_i I_i \int (\nabla \times \mathbf{A}_i) \cdot d\mathbf{S}_i = \frac{1}{2c} \sum_i I_i \int \mathbf{B}_i \cdot d\mathbf{S}_i = \frac{1}{2c} \sum_i I_i \Phi_i,
 \end{aligned}$$

where $\Phi_i = \int \mathbf{B}_i \cdot d\mathbf{S}_i$ is the flux of magnetic induction through the i th conductor.

Let us find the coefficient of self-inductance for certain simple systems.

As the first example we shall consider the coefficient of self-inductance of a toroidal solenoid. The field inside the solenoid is given by formula (24.18). The energy of the field is equal to

$$\frac{\mu H^2 V}{8\pi} = \frac{\mu n^2 I^2 S \cdot 2\pi D}{2\pi c^2 D^2} = \frac{LI^2}{2},$$

hence

$$L = \frac{2\mu n^2 S}{c^2 D}.$$

Let us now consider the coefficient of self-inductance of two coaxial cylinders along which identical and oppositely directed currents of intensity I are flowing. Let the radii of the cylinders be R_1 and R_2 , their length l , and the thickness of the walls negligibly small. It is obvious that the magnetic field inside the small cylinder and outside the large cylinder is equal to zero. The field in the gap between the cylinders has the strength

$$H = \frac{2I}{cr}.$$

The energy of the field is equal to

$$\int \frac{\mu H^2}{8\pi} dV = \frac{\mu I^2 l}{c^2} \int_{R_1}^{R_2} \frac{r dr}{r^2} = \frac{\mu I^2 l}{c^2} \ln \frac{R_2}{R_1}.$$

Hence

$$L = \frac{2\mu l}{c^2} \ln \frac{R_2}{R_1}.$$

§26. Coefficients of self-inductance and mutual inductance for non-linear conductors

In calculating coefficients of self-inductance and mutual inductance for non-linear conductors we shall confine ourselves to the case where the magnetic permeability of the conductors and the medium is close to unity (i.e.

where there are no ferromagnetic materials in the system). In this case it can be assumed approximately that μ has the same value μ_0 over all space. In other words, we shall assume the entire system to be homogeneous in its magnetic properties.

For brevity of notation we shall assume that the system is made up of two non-linear conductors. We shall assume the geometrical form and the current density distribution inside the conductors to be invariable in time, and the conductors to be at rest and non-deformable.

We divide the two non-linear conductors into a set of current tubes. The possibility of such a division follows from the solenoidal character of currents in quasistationary fields. For each tube dS_α one can write the relation (24.6), i.e.

$$\oint \mathbf{j}_\alpha \cdot \frac{d\mathbf{l}_\alpha}{\sigma_\alpha} = \oint \mathbf{E}_\alpha \cdot d\mathbf{l}_\alpha + \oint \mathbf{E}_\alpha^{\text{imp}} \cdot d\mathbf{l}_\alpha. \quad (26.1)$$

The relations (16.5) – (16.7), which have been obtained for any solenoidal current, are valid for quasistationary currents. Multiplying both sides of the eq. (26.1) by $I^{-1}dI_\alpha$, a quantity which is constant along the conductor, and integrating over the entire section of the conductor, we have

$$\int_S \frac{dI_\alpha}{I} \oint \mathbf{j}_\alpha \cdot \frac{d\mathbf{l}_\alpha}{\sigma_\alpha} = \int_S \frac{dI_\alpha}{I} \oint \mathbf{E}_\alpha \cdot d\mathbf{l}_\alpha + \int_S \frac{dI_\alpha}{I} \oint \mathbf{E}_\alpha^{\text{imp}} \cdot d\mathbf{l}_\alpha. \quad (26.2)$$

We transform the integrals in (26.2) for example for the first conductor in the following way:

$$\begin{aligned} \int_{S_1} \frac{dI_\alpha^{(1)}}{I_1} \oint \mathbf{j}_1 \cdot \frac{d\mathbf{l}_1}{\sigma_1} &= \iint \frac{dI_\alpha^{(1)}}{I_1} \mathbf{j}_1 \cdot \frac{d\mathbf{l}_1}{\sigma_1} = \int_{S_1} \oint \psi_1 \frac{dS_\alpha^{(1)}}{S_1} \mathbf{j}_1 \cdot \frac{d\mathbf{l}_1}{\sigma_1} = \\ &= \iint \frac{I_1 \psi_1^2 dS_\alpha^{(1)} d\mathbf{l}_1}{\sigma_1 S_1^2} = I_1 \int_{S_1} \oint \frac{\psi_1^2 dS_\alpha^{(1)} d\mathbf{l}_1}{\sigma_1 S_1^2} = I_1 R_1, \end{aligned} \quad (26.3)$$

where R_1 is the total ohmic resistance of the first conductor. Here we have made use of (16.7) and (16.8). For a linear conductor $\psi = 1$ and the expression for R is the same as (15.4). Further, on the basis of (23.7) and (24.10) we have

$$\begin{aligned}
 \int_{S_1} \frac{dI_{\alpha}^{(1)}}{I} \oint \mathbf{E}_1 \cdot d\mathbf{l}_1 &= -\frac{1}{c} \int_{S_1} \frac{dI_{\alpha}^{(1)}}{I} \frac{d\Phi_1}{dt} = -\frac{1}{c} \frac{d}{dt} \oint_S \frac{dI_{\alpha}}{I} \mathbf{A}_1 \cdot d\mathbf{l}_1 = \\
 &= -\frac{1}{c} \frac{d}{dt} \oint_S \mathbf{j}_1 \cdot \frac{\mathbf{A}_1 dI_1 dS_{\alpha}}{I} = -\frac{1}{c} \frac{d}{dt} \frac{1}{I} \int \mathbf{j}_1 \cdot \mathbf{A}_1 dv, \quad (26.4)
 \end{aligned}$$

Here \mathbf{A}_1 is the vector potential of the magnetic field at a point in the volume element dv .

We shall now, for definiteness, consider the first conductor. Since by assumption the conductors and the medium are homogeneous in respect to their magnetic properties, we can make use of the expression (17.3) for \mathbf{A}_1 , which for our purposes is conveniently written in the form

$$\mathbf{A}_1 = \frac{\mu_0}{c} \int \mathbf{j}' \frac{dV'}{r} = \frac{\mu_0}{c} \int \mathbf{j}'_1 \frac{dv'_1}{r'_{11}} + \frac{\mu_0}{c} \int \mathbf{j}_2 \frac{dv_2}{r_{12}}. \quad (26.5)$$

The first term gives the vector potential produced by the currents in the volume of the first conductor, the second term gives the vector potential produced by the currents in the volume of the second conductor, r'_{11} denotes the distance between the points to which the element dv'_1 and the element $d\mathbf{l}_1$ are referred respectively, and r_{12} has an analogous meaning.

Substituting the value of \mathbf{A}_1 into (26.4), we have

$$\begin{aligned}
 \int_{S_1} \frac{dI_{\alpha}^{(1)}}{I_1} \oint \mathbf{E}_1 \cdot d\mathbf{l}_1 &= \\
 &= -\frac{\mu_0}{c^2} \frac{d}{dt} \frac{1}{I_1} \iint \frac{\mathbf{j}_1 \cdot \mathbf{j}'_1 dv_1 dv'_1}{r'_{11}} - \frac{\mu_0}{c^2} \frac{d}{dt} \frac{1}{I_1} \iint \frac{\mathbf{j}_1 \cdot \mathbf{j}_2 dv_1 dv_2}{r_{12}}. \quad (26.6)
 \end{aligned}$$

We now introduce the coefficients of self-inductance and mutual inductance, defining them by the formulae:

$$L_{11} = \frac{1}{I_1^2} \frac{\mu_0}{c^2} \iint \frac{\mathbf{j}'_1 \cdot \mathbf{j}_1 dv'_1 dv_1}{r'_{11}}, \quad (26.7)$$

$$L_{12} = \frac{1}{I_1 I_2} \frac{\mu_0}{c^2} \iint \frac{\mathbf{j}_1 \cdot \mathbf{j}_2 dv_1 dv_2}{r_{12}}, \quad (26.8)$$

$$L_{21} = L_{12} . \quad (26.9)$$

It is easily seen that L_{11} and L_{12} do not depend on the current intensities I_1 and I_2 . Indeed, from (16.8) we have

$$L_{11} = \frac{\mu_0}{c^2} \iint \frac{\psi_1 \psi'_1}{S_1 S'_1} \frac{\mathbf{l}_1 \cdot \mathbf{l}'_1}{r'_{11}} dv'_1 dv_1 , \quad (26.10)$$

$$L_{12} = \frac{\mu_0}{c^2} \iint \frac{\psi_1 \psi_2}{S_1 S_2} \frac{\mathbf{l}_1 \cdot \mathbf{l}_2}{r_{12}} dv_1 dv_2 . \quad (26.11)$$

Then, finally,

$$\int_{S_1} \frac{dI_\alpha}{I_1} \oint \mathbf{E}_1 \cdot d\mathbf{l}_1 = -L_{11} \frac{dI_1}{dt} - L_{12} \frac{dI_2}{dt} \quad (26.12)$$

and analogously for the second conductor.

In the integral $\int I^{-1} dI_\alpha \oint \mathbf{E}^{\text{imp}} \cdot d\mathbf{l}$ one can assume that the impressed e.m.f. is the same over the whole section of the conductor and write

$$\int \frac{dI'_\alpha}{I_1} \oint \mathbf{E}_1^{\text{imp}} \cdot d\mathbf{l}_1 = \mathcal{E}_1 . \quad (26.13)$$

Substituting (26.3), (26.12) and (26.13) into (26.2), we find finally

$$I_1 R_1 = -L_{11} \frac{dI_1}{dt} - L_{12} \frac{dI_2}{dt} + \mathcal{E}_1 \quad (26.14)$$

and analogously

$$I_2 R_2 = -L_{22} \frac{dI_2}{dt} - L_{21} \frac{dI_1}{dt} + \mathcal{E}_2 . \quad (26.15)$$

Eqs. (26.14) and (26.15) are the same as eqs. (24.17) for linear currents. A difference between them lies in the fact that for non-linear conductors finding the resistances and coefficients of inductance is a very complex problem. All these quantities involve current distributions over the cross-sections of the conductors. Therefore the practical importance of the equations obtained is not great. To make up for it we have obtained a definition of coefficients of inductance (26.7) – (26.9) which is not based (in contrast to (24.12) and (24.13)) on the assumption of linearity of the conductors.

In conclusion we stress that, as is easily seen, the coefficient of mutual inductance defined by formula (26.8) for a linear conductor ($\psi_1 = \psi_2 = 1$) is the same as (24.12).

§27. Lagrange's equations for a system of quasistationary currents

Up to now we have assumed that the relative position of the current carrying conductors is given. In practice one often has to consider the more general case of moving conductors or conductors changing their shape, relative position etc., which are placed in an electromagnetic field. The capacitances, self-inductances, mutual inductances and other quantities, which we have before assumed to be constant, turn out in this case to be functions of certain parameters q_i which characterize the configuration of the system. In considering such systems it is very convenient to make use of Lagrange's method in which, as will be seen from what follows, the electromagnetic and mechanical quantities characterizing the system come in as formally equivalent quantities. It turns out that the quoted parameters q_i , as well as the values of charges Q_i in the conductors, can be chosen as generalized coordinates.

Let us work out the Lagrange equations characterizing the system in generalized coordinates q_i and Q_i . The generalized velocities corresponding to the charges Q_i are the quantities $\dot{Q}_i = I_i$, i.e. currents flowing in the conductors. If the energy of the magnetic field is included in the kinetic energy of the system, then the total kinetic energy is equal to

$$\begin{aligned} T = T_{\text{mech}} + T_{\text{magn}} &= \frac{1}{2} \sum_{i,k} L_{ik}(q_i, t) I_i(t) I_k(t) + T_{\text{mech}} = \\ &= \frac{1}{2} \sum_{i,k} L_{ik} \dot{Q}_i \dot{Q}_k + T_{\text{mech}}, \end{aligned} \quad (27.1)$$

where T_{mech} is the kinetic energy of the mechanical motion of the conductors.

Analogously, we assume the potential energy of the system to be made up of the energy of the electric field (i.e. the energy of the capacitances present in the system) and the mechanical potential energy:

$$U = \sum \frac{Q_i^2(t)}{2C_i(q_i, t)} + U_{\text{mech}}. \quad (27.2)$$

Then the Lagrange function will have the form

$$L = \frac{1}{2} \sum L_{ik} \dot{Q}_i \dot{Q}_k - \frac{1}{2} \sum \frac{Q_i^2}{C_i} + (T_{\text{mech}} - U_{\text{mech}}). \quad (27.3)$$

We also introduce the dissipative function

$$F = \frac{1}{2} \sum R_i(q_i, t) \dot{Q}_i^2 \quad (27.4)$$

and arbitrary "external" forces acting on the system — impressed e.m.f. \mathcal{E}_i . We recall that external forces F_{ext} in the Lagrange equations represent forces which depend not only on the generalized coordinates but also on other parameters bearing no relation to the given system. These forces are defined by the usual relation $\delta A = F_{\text{ext}} \delta q$, where δA is the virtual work done in the virtual displacement δq .

Then the Lagrange equations for the generalized coordinates Q_i assume the form

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{Q}_i} - \frac{\partial L}{\partial Q_i} = - \frac{\partial F}{\partial \dot{Q}_i} + F_{\text{ext}}. \quad (27.5)$$

Upon substituting L into (27.5) and assuming that the current loops are motionless and non-deformable (constant inductances and capacitances) we obtain

$$\sum L_{ik} \ddot{Q}_k + \frac{Q_i}{C_i} = - R_i \dot{Q}_i + \mathcal{E}_i, \quad (27.6)$$

which shows that the expressions for the kinetic and potential energies are appropriately chosen.

From the Lagrange equation (27.6) there follows the momentum conservation law and the energy conservation law. For a system which is not acted upon by any external forces (i.e. $F_{\text{ext}} = 0$) and in which no energy dissipation occurs (i.e. $F = 0$) eq. (27.5) assumes the form

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{Q}_i} - \frac{\partial L}{\partial Q_i} = 0.$$

If there is no capacitance in the i th current loop, then the corresponding coordinate is cyclic and the following equality holds for it:

$$\frac{\partial L}{\partial Q_i} = 0.$$

In this case the corresponding generalized momentum is conserved:

$$\frac{\partial L}{\partial \dot{Q}_i} = \frac{\partial L}{\partial I_i} = \sum L_{ik} I_k = \text{const}.$$

Let us now write the energy conservation law for a system of conductors at rest. According to general rules, the energy of the system is

$$E = \sum \frac{\partial L}{\partial \dot{Q}_i} \dot{Q}_i - L = \frac{1}{2} \sum L_{ik} \dot{Q}_i \dot{Q}_k + \frac{1}{2} \sum \frac{Q_i^2}{C_i}. \quad (27.7)$$

In the presence of dissipative processes

$$\frac{dE}{dt} = -2F + \sum F_{\text{ext}} \dot{Q}_i.$$

Here $2F$ is the energy dissipated, and $\sum F_{\text{ext}} \dot{Q}_i$ is the work done by external forces per unit time.

Substituting the values of E and F , we write the energy conservation law in the form

$$\frac{d}{dt} \left(\frac{1}{2} \sum L_{ik} \dot{Q}_i \dot{Q}_k + \frac{1}{2} \sum \frac{Q_i^2}{C_i} \right) = - \sum R_i \dot{Q}_i^2 + \sum \mathcal{E}_i \dot{Q}_i. \quad (27.8)$$

Rewriting it in the form

$$\frac{d}{dt} (T_{\text{magn}} + U_{\text{el}}) = \sum (\mathcal{E}_i I_i - R_i I_i^2), \quad (27.9)$$

we see that in a closed system which is not acted upon by any external forces ($\mathcal{E}_i = 0$) and in which there are no dissipative forces ($R_i \neq 0$) the energy of the system is conserved:

$$T_{\text{magn}} + U_{\text{el}} = 0. \quad (27.10)$$

In the case where $\mathcal{E}_i \neq 0$, $R_i \neq 0$, formula (27.9) shows that the difference between the work done by the impressed e.m.f. and the Joule heat released goes into an increase in the energy of the magnetic and electric fields as well as into the increase in the mechanical energy of the system in the case of a system of moving current contours.

In the particular case of one motionless current loop, (27.9) can be rewritten in the form

$$\frac{d}{dt} \left(\frac{LI^2}{2} + \frac{Q^2}{2C} \right) = \mathcal{E}I - RI^2.$$

The relations obtained are valid for a system of loops which are coupled with each other inductively (i.e. in the presence of coefficients L_{ik} differing from zero) as well as for a system of branched loops. In the latter case Kirchhoff's law allows one to reduce the number of independent currents I_i or coordinates Q_i .

In conclusion we note that the results of this section justify to a certain degree the term "impressed e.m.f.", since this quantity indeed plays the role of a generalized force acting on the system.

We apply the results obtained to a single alternating current circuit of resistance, inductance and capacitance in series (the so-called RLC -series circuit). Eq. (27.6) assumes the form

$$L \frac{d^2 Q}{dt^2} + \frac{Q}{C} = -R\dot{Q} + \mathcal{E}(t). \quad (27.11)$$

If the source of the alternating current (an impressed e.m.f.) represents a harmonic function of time

$$\mathcal{E}(t) = \mathcal{E}_0 e^{i\omega t},$$

then the particular solution of (27.11) has the form

$$Q = Q_0 e^{i\omega t}.$$

We are interested in the dependence of the current on time, which also can be written in the form

$$I = I_0 e^{i\omega t}.$$

Substituting this into (27.11), we obtain

$$I_0 = \mathcal{E}_0 / Z^*,$$

where the quantity Z^* , called the impedance, is equal to

$$Z^* = R + i \left(\omega L - \frac{1}{\omega C} \right). \quad (27.12)$$

Passing over from the complex expression to the real expression for the current, we find

$$I = \frac{\mathcal{E}_0 \cos(\omega t - \varphi)}{[R^2 + (\omega L - 1/\omega C)^2]^{\frac{1}{2}}}, \quad (27.13)$$

$$\varphi = \arctan \frac{(\omega L - 1/\omega C)}{R}.$$

We see that forced current oscillations with frequency ω , which are shifted in phase by an angle φ with respect to the impressed e.m.f., arise in the circuit.

Besides the forced oscillations there can exist free or natural oscillations. If the impressed e.m.f. is $\mathcal{E} = 0$, then from (27.11) we easily find

$$I = I_0 e^{i\omega_0 t},$$

where the frequency of the natural oscillations is

$$\omega_0 = i \frac{R}{2C} \pm \left(\frac{1}{LC} - \left(\frac{R}{2L} \right)^2 \right)^{\frac{1}{2}}. \quad (27.14)$$

For $R/2C < (CL)^{-\frac{1}{2}}$ damped oscillations occur, the damping being characterized by the time $\tau = 2C/R$. For $R/2C > (CL)^{-\frac{1}{2}}$ the discharge has an aperiodic character.

If the system contains several loops which are connected with each other by corresponding coefficients of mutual inductance, then the system of eqs. (27.6) can be solved according to the general rules of solution of a system of linear equations with constant coefficients. For a detailed analysis of such problems, which is of special interest in electrical engineering, we refer the reader to specific literature *.

* See, for example, W.R.Smythe, *Static and dynamic electricity* (McGraw-Hill, New York, 1950).

§28. The generalized ponderomotive forces in a system with moving current loops

In the preceding section we have confined ourselves to the case of motionless loops. We shall now consider moving current loops.

We shall find, first of all, an expression for the generalized force corresponding to the mechanical generalized coordinate q_i characterizing the spatial configuration of the i th current loop. By definition

$$F_{q_i} = \frac{\partial L}{\partial q_i}.$$

It can usually be assumed that the capacitances which are present in the system do not change when the conductors are moving. Then substituting L from (27.3), we find

$$F_{q_i} = \frac{\partial L}{\partial q_i} = \frac{\partial}{\partial q_i} T_{\text{magn}} = \frac{\partial}{\partial q_i} \frac{1}{2} \sum L_{ik}(q_i, q_k) I_i I_k. \quad (28.1)$$

We see that the energy of the magnetic field plays a double role: with respect to the coordinates of electromagnetic character, Q_i , it represents a kinetic energy, while with respect to the space coordinates q_i it represents a potential energy taken with the opposite sign. Ponderomotive forces act in a direction which corresponds to an increase of the magnetic energy of the field.

We shall apply formula (28.1) to certain practical cases.

Let us first consider mechanical forces acting on a single current loop. Let q be the generalized coordinate characterizing the size of the loop. The force acting on the loop on the part of its own magnetic field is

$$F_q = \frac{I^2}{2} \frac{\partial L_{ii}}{\partial q}, \quad (28.1')$$

The force acting on the current loop increases with increasing $\partial L_{ii}/\partial q$. Since the self-inductance increases with increasing size of the loop, this means that the magnetic field of the current loop tends to deform the conductor in such a way that its size would increase.

For a system of conductors we shall first consider one moving loop whose position is characterized by the coordinate q_α . We assume that the configuration and currents in all other loops in the system are given. Then the force acting on the moving loop is

$$\begin{aligned}
 F_{q_\alpha} &= \frac{\partial}{\partial q_\alpha} \frac{1}{2} \sum_{i,k} L_{ik}(q_i, q_k) I_i I_k = \\
 &= \frac{1}{2} \sum_k \frac{\partial L_{\alpha k}}{\partial q_\alpha} I_\alpha I_k + \frac{1}{2} \sum_i \frac{\partial L_{i\alpha}}{\partial q_\alpha} I_i I_\alpha = \\
 &= I_\alpha \sum_k \frac{\partial L_{\alpha k}}{\partial q_\alpha} I_k = \frac{I_\alpha}{c} \frac{\partial \Phi_\alpha}{\partial q_\alpha}, \quad (28.2)
 \end{aligned}$$

The fixed positions in space of all the loops except the α th correspond to a given value of the external field. In this derivation we have made use of the equality (24.9): the definition of the flux of induction.

If, in particular, the generalized coordinate q_α represents the angle θ which characterizes the position of a flat current loop in an external field, then the generalized force F_θ represents the torque acting on the loop:

$$M = \frac{I}{c} \frac{\partial \Phi}{\partial \theta}.$$

The flux of induction through a non-deformable flat current loop in an external field \mathbf{B}_0 is equal to

$$\Phi = B_0 S \cos \theta,$$

where S is the area of the contour, and θ is the angle between \mathbf{B}_0 and the normal to the plane. Hence

$$M = -\frac{IS}{c} B_0 \sin \theta. \quad (28.3)$$

We note that the generalized force (28.2) reduces, in the case of one current loop in an external magnetic field, to the Lorentz force. Indeed, writing for $\delta\Phi$

$$\delta\Phi = \mathbf{B} \cdot (\delta\mathbf{q} \times \delta\mathbf{l}) = (\delta\mathbf{l} \times \mathbf{B}) \cdot \delta\mathbf{q},$$

we have

$$d\mathbf{F} = c^{-1} I(d\mathbf{l} \times \mathbf{B}) = c^{-1} (d\mathbf{l} \times \mathbf{B})(\mathbf{j} \cdot d\mathbf{S}) = c^{-1} (\mathbf{j} \times \mathbf{B}) dV, \quad (28.4)$$

where $dV = d\mathbf{S} \cdot d\mathbf{l}$ is a volume element of the current carrying conductor. The force (28.4) is the averaged Lorentz force in which the mean magnetic field in the medium \mathbf{B} appears instead of the field in vacuum \mathbf{H} .

Let us now consider a system of two current carrying conductors and find the force of interaction between them. In this case one has to choose the distance r_{12} between the elements $d\mathbf{l}_1$ and $d\mathbf{l}_2$ of the two conductors as the generalized coordinate q .

From the definition of the coefficient of self-inductance (24.12), (24.11) and (28.2) we have

$$\mathbf{F} = I_1 I_2 \frac{\mu}{c^2} \frac{\partial}{\partial r_{12}} \oint \oint \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r_{12}} = - \frac{\mu I_1 I_2}{c^2} \oint \oint \frac{r_{12} d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r_{12}^3}. \quad (28.5)$$

In particular, let the two currents flow in the same direction in parallel linear conductors. Then $d\mathbf{l}_1 \parallel d\mathbf{l}_2$ and the force \mathbf{F} corresponds to the attraction between the two current carrying conductors.

On the other hand, in the case of currents flowing in opposite directions, a repulsion arises between the conductors.

We see that two currents in the same direction tend to approach each other and thus intensify their common magnetic field. It is in this respect that they differ from static charges of the same sign, which tend to separate and thus weaken their electrostatic field.

We shall consider in addition the problem of the work which is done by the force displacing a current loop. According to (28.2), this work is equal to

$$\delta W = F_{q_\alpha} \delta q_\alpha = c^{-1} I_\alpha \delta \Phi_\alpha.$$

At first sight we arrive at a paradoxical result: the work δW is done by the force due to the magnetic field acting on the charges. This force, however, is perpendicular to the velocity of the charges and cannot do any work on them. In fact, however, in writing the formula for the work we have not taken into account induction phenomena which occur in a conductor moving in a magnetic field. As the conductor moves, an e.m.f. \mathbf{E}^{ind} is induced in it, and work $\delta W'$ equal to

$$\delta W' = \delta t \int \mathbf{j} \cdot \mathbf{E}^{\text{ind}} dV = I \delta t \oint \mathbf{E}^{\text{ind}} \cdot d\mathbf{l} = - c^{-1} I_\alpha \delta \Phi_\alpha$$

is done on the charges. The total work done by the magnetic field on the current loop is

$$\delta W + \delta W' = 0,$$

as was to be expected.

This result is of a general character and can also be applied to non-linear conductors.

When we have spoken about the work of magnetization and the change in the thermodynamic potentials of the system which is associated with it, we have meant by this the work done on the currents by the electric field induced when the magnetic field is switched on. We write it in another, more general form:

$$\begin{aligned}\delta W' &= \delta t \int \mathbf{j} \cdot \mathbf{E} dV = \delta t \frac{c}{4\pi} \int \mathbf{E} \cdot (\nabla \times \mathbf{H}) dV = \\ &= -\frac{c}{4\pi} \delta t \int \nabla \cdot (\mathbf{E} \times \mathbf{H}) dV + \frac{c}{4\pi} \delta t \int \mathbf{H} \cdot (\nabla \times \mathbf{E}) dV = \\ &= -\frac{c}{4\pi} \delta t \oint (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{S} - \frac{1}{4\pi} \delta t \int \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} dV = \\ &= -\frac{1}{4\pi} \int \mathbf{H} \cdot \left(\delta t \frac{\partial \mathbf{B}}{\partial t} \right) dV = -\frac{1}{4\pi} \int \mathbf{H} \cdot \delta \mathbf{B} dV.\end{aligned}$$

Hence it follows that the body which is magnetized gets an additional energy $-\delta W'$.

Therefore the free energy per unit volume can be written in the form

$$dF = dF_0 + \frac{1}{4\pi} \mathbf{H} \cdot d\mathbf{B}.$$

We define the thermodynamic potential G_1 as

$$G_1 = F + TS - \mathbf{B} \cdot \mathbf{H}/4\pi.$$

For $\mathbf{B} = \mu \mathbf{H}$ we have

$$G_1 = F + TS - \mu H^2/8\pi.$$

For dG we have

$$dG = dG_1 - \mathbf{B} \cdot d\mathbf{H}/4\pi.$$

Hence

$$\mathbf{B} = -4\pi(\partial G/\partial \mathbf{H})_{T,p}.$$

To obtain (20.3) use can be made of the last formula, expressing \mathbf{B} in terms of \mathbf{M} :

$$\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M} = -4\pi(\partial G/\partial \mathbf{H})_{T,p}.$$

Integrating with respect to \mathbf{H} and assuming \mathbf{M} to be the independent variable, we obtain

$$G = G_1 - \mathbf{M} \cdot \mathbf{H} - H^2/8\pi.$$

The quantity $\int (H^2/8\pi)dV$ represents the energy of the external field in the volume of the body.

In conclusion we shall consider an example which illustrates in an obvious way the merits of the Lagrange method.

Let there be two loops: the first carrying a current I_1 and rotating inside the second under the action of a force F , the second loop being at rest. A current, I_2 , is induced in the second loop. In both loops there is no capacitance. It is clear that such an arrangement represents an alternating-current machine. The first loop is called the rotor, and the second loop is called the stator. The Lagrange function of the system has the form

$$L = \frac{1}{2}L_{11}I_1^2 + L_{12}I_1I_2 + \frac{1}{2}L_{22}I_2^2 + \frac{1}{2}I_0(d\alpha/dt)^2 - U_{\text{mech}},$$

where α is the angle of rotation, and I_0 is the moment of inertia of the rotor. The coefficient of mutual inductance depends on the orientation of the rotor and stator, i.e. on the angle α .

The equation of motion for the generalized coordinate, the angle α , has the form

$$I_0\ddot{\alpha} - I_1I_2 \frac{dL_{12}}{d\alpha} = M_0, \quad (28.6)$$

where $M_0 = -\partial U_{\text{mech}}/\partial \alpha$ is the moment of the rotational force.

The equation of motion for the generalized coordinate, the charge Q_2 , i.e. the equation of the current in the stator, according to (27.5), has the form

$$\frac{d}{dt} \frac{\partial L}{\partial I_2} = -R_2 I_2,$$

or

$$\frac{d}{dt} (L_{22} I_2 + L_{12} I_1) = -R_2 I_2.$$

Hence

$$L_{22} \frac{dI_2}{dt} + R_2 I_2 + L_{12} \frac{dI_1}{dt} + I_1 \frac{dL_{12}}{dt} = 0. \quad (28.7)$$

The state of the system is to be found from the simultaneous solution of eqs. (28.6) and (28.7).

Assuming the resistance R_2 to be very large (a large load resistance in the circuit of the stator), one can write approximately

$$I_2 \approx -\frac{I_1}{R_2} \frac{dL_{12}}{dt} = -\frac{I_1}{R_2} \frac{dL_{12}}{d\alpha} \frac{d\alpha}{dt}. \quad (28.8)$$

We have neglected small terms in (28.7) which do not contain R_2 . Substituting (28.8) into (28.6), we obtain

$$I_0 \frac{d^2 \alpha}{dt^2} + \frac{I_1^2}{R_2} \left(\frac{dL_{12}}{d\alpha} \right)^2 \frac{d\alpha}{dt} = M_0.$$

For the quasistationary state one can disregard the term with the angular acceleration $d^2 \alpha / dt^2$, since it is small. Then for the angular velocity of rotation we find

$$\frac{d\alpha}{dt} = \frac{M_0 R_2}{I_1^2} \left(\frac{dL_{12}}{d\alpha} \right)^{-2} = \text{const}. \quad (28.9)$$

The expressions (28.8) and (28.9) give the solution of the problem. In practice one usually makes use of more complex schemes, for example machines with self-excitation in which the induced current I_2 is introduced into the first

We have not considered the important properties of real machines associated with the presence of magnetizable cores. The example discussed illustrates only the merits of the Lagrange method in the case of systems in which the mechanical motion and currents are directly connected with each other.

§ 29. Fluctuations in conductors and the Nyquist formula

As a result of fluctuation processes in an electric circuit current fluctuations arise which in practice are called noise. Physically the appearance of fluctuation currents in a conductor (in the absence of impressed e.m.f.) is associated with the fluctuations of the number of electrons moving in a particular direction. In the presence of an impressed e.m.f., fluctuation currents are superimposed upon the stationary or quasistationary current.

Current fluctuations in a radio set are of very great importance in radio engineering. The noise background determines the limit of sensitivity of reception of a single signal. A further increase in sensitivity can be achieved only by repeated measurements.

We shall consider the theory of fluctuations in an electric circuit with inductance L and ohmic resistance R . Fluctuation processes in the circuit can be characterized by a random fluctuating e.m.f. $\mathcal{E}(t)$. The variations of the e.m.f. $\mathcal{E}(t)$ occur in a time which is very small in comparison with the relaxation time $T = L/R$ of the circuit.

It is natural to try to describe the processes occurring in the circuit by means of the generalized Ohm's law

$$L \frac{di}{dt} + Ri = \mathcal{E}(t). \quad (29.1)$$

Eq. (29.1) is based on the rather natural assumption that between the random current in a circuit and the random e.m.f. producing it there is the same relation as between an ordinary current and the ordinary e.m.f. The current relaxation is characterized by a constant resistance R . It should be stressed that at high frequencies the resistance R (or $\sigma = R^{-1}$; see § 14) turns out to be a function of the frequency. This fact does not however affect the general results of the theory.

The random e.m.f. $\mathcal{E}(t)$ has, obviously, the following properties:

$$\overline{\mathcal{E}(t)} = 0 \quad \overline{[\mathcal{E}(t)]^2} \neq 0, \quad (29.2)$$

We integrate the linear equation (29.1) with respect to time. We then have for the random current:

$$i(t) = i_0 e^{-t/T} + e^{-t/T} \int_0^t e^{t'/T} \mathcal{E}(t') dt'. \quad (29.3)$$

It should be noted that the integration of eq. (29.1) with the random function on the right-hand side requires some justification from the purely mathematical point of view. For the details of this problem we refer the reader to more specialized literature*.

By means of formula (5.3) of Part III one can work out the correlation function:

$$\begin{aligned} \langle i(t) i(t+\tau) \rangle &= \left\langle \left[i_0 e^{-t/T} + e^{-t/T} \int_0^t e^{t'/T} \mathcal{E}(t') dt' \right] \right. \\ &\quad \times \left. \left[i_0 e^{-(t+\tau)/T} + e^{-(t+\tau)/T} \int_0^{t+\tau} e^{t''/T} \mathcal{E}(t'') dt'' \right] \right\rangle = \\ &= i_0^2 e^{-(2t+\tau)/T} \\ &\quad + e^{-(2t+\tau)/T} \int_0^t \int_0^{t+\tau} e^{(t'+t'')/T} \langle \mathcal{E}(t') \mathcal{E}(t'') \rangle dt' dt''. \end{aligned} \quad (29.4)$$

Terms containing the random function $\mathcal{E}(t)$ in the first power reduce to zero in averaging.

The double integral in (29.4) can be calculated in the following way. Introducing new variables $x = t' + t''$, $y = t' - t''$, we have

$$I(t, \tau) = \frac{1}{2} e^{-(2t+\tau)/T} \int_0^{2t} e^{x/T} dx \int_{-t}^t \langle \mathcal{E}(\frac{1}{2}x - \frac{1}{2}y) \mathcal{E}(\frac{1}{2}x + \frac{1}{2}y) \rangle dy.$$

The correlation function $\langle \mathcal{E}(\frac{1}{2}x - \frac{1}{2}y) \mathcal{E}(\frac{1}{2}x + \frac{1}{2}y) \rangle$, cannot, according to (5.4) of Part III, depend on the choice of the variable x . Assuming that $x = y$, we have

* See, for example, S.Chandrasekhar, *Stochastic problems in physics and astronomy*,

$$\langle \mathcal{E}(\frac{1}{2}x - \frac{1}{2}y) \mathcal{E}(\frac{1}{2}x + \frac{1}{2}y) \rangle \equiv \langle \mathcal{E}(0) \mathcal{E}(y) \rangle .$$

Since the correlation function rapidly decreases with increasing y , the range of integration can be extended to infinity, so that

$$\begin{aligned} I(t, \tau) &= \frac{1}{2} e^{-(2t+\tau)/T} \int_0^{2t} e^{x/T} dx \int_{-\infty}^{\infty} \langle \mathcal{E}(0) \mathcal{E}(y) \rangle dy = \\ &= \frac{1}{2} T e^{-\tau/T} (1 - e^{-2t/T}) \int_{-\infty}^{\infty} \langle \mathcal{E}(0) \mathcal{E}(y) \rangle dy . \end{aligned}$$

Finally we find

$$\langle i(t) i(t+\tau) \rangle = i_0^2 e^{-(2t+\tau)/T} + \frac{1}{2} T e^{-\tau/T} (1 - e^{-2t/T}) \int_{-\infty}^{\infty} \langle \mathcal{E}(0) \mathcal{E}(y) \rangle dy . \quad (29.5)$$

Assuming $t = 0$, we obtain from (29.5) the correlation function for the current

$$\langle i(0) i(\tau) \rangle = i_0^2 e^{-\tau/T} . \quad (29.6)$$

Making use of the law of equipartition, one can write the mean value of the fluctuation current in the circuit $\frac{1}{2} L i_0^2 = \frac{1}{2} k T \equiv \frac{1}{2} \theta$ so that finally the autocorrelation function for the current assumes the form

$$\langle i(0) i(\tau) \rangle = \theta L^{-1} e^{-|\tau|/T} . \quad (29.7)$$

On the other hand, assuming $\tau = 0$, we find

$$\langle i(t)^2 \rangle = i_0^2 e^{-2t/T} + \frac{1}{2} T (1 - e^{-2t/T}) \int_{-\infty}^{\infty} \langle \mathcal{E}(0) \mathcal{E}(y) \rangle dy .$$

For large values of t , $t \gg T$, a statistical equilibrium must be established in the system. Dropping as small the quantities with $e^{-2t/T}$, we find

$$\langle i(t)^2 \rangle = \theta L^{-1} = \frac{1}{2} T \int_{-\infty}^{\infty} \langle \mathcal{E}(0) \mathcal{E}(y) \rangle dy .$$

Hence we find, substituting $T = LR^{-1}$:

$$R = \frac{1}{2} L^2 \theta^{-1} \int_{-\infty}^{\infty} \langle \mathcal{E}(0) \mathcal{E}(y) \rangle dy . \quad (29.8)$$

Formula (29.8) relates the autocorrelation function of the random e.m.f. and the resistance of the circuit. This relation has a deep meaning: it connects a characteristic of reversible random processes, the correlation function $\langle \mathcal{E}(0) \mathcal{E}(y) \rangle$, with a characteristic of the irreversible dissipative process, the relaxation time T or the resistance R

$$T^{-1} = \frac{1}{2} L \theta^{-1} \int_{-\infty}^{\infty} \langle \mathcal{E}(0) \mathcal{E}(y) \rangle dy . \quad (29.9)$$

If one makes use of the Wiener–Khinchin formula (5.16) of Part III, then one can express the correlation function in terms of the spectral density. Then we have

$$\frac{1}{2} \pi^{-1} \int_{-\infty}^{\infty} \langle \mathcal{E}(0) \mathcal{E}(y) \rangle dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i\omega y} g(\omega) d\omega dy = g(0) , \quad (29.10)$$

and, consequently,

$$R = \frac{1}{2} L^2 \theta^{-1} g(0) . \quad (29.11)$$

This formula is a particular case of an important fluctuation–dissipation theorem establishing a connection between the characteristics of fluctuation processes and dissipative processes.

In formula (29.1) we have restricted ourselves to the case where the resistance can be assumed to be independent of the frequency. The problem of the connection between fluctuation processes and irreversible processes will be considered in detail in Part VI which is devoted to physical kinetics.

We return to the autocorrelation function for the current and, making use of the Wiener–Khinchin theorem, we find the spectral density of random currents in the circuit. Namely, according to (5.16) of Part III and (29.11), we can write

$$g_I(\omega) = \frac{1}{2} \pi^{-1} \int_{-\infty}^{\infty} \langle i(0) i(\tau) \rangle e^{-i\omega \tau} d\tau , \quad (29.12)$$

where the spectral density of the current $g_i(\omega)$ is defined by the formula

$$\overline{i^2} = \int_0^\infty g_i(\omega) d\omega. \quad (29.13)$$

Substituting the expression for $\langle i(0)i(\tau) \rangle$ from (29.7) into (29.12), we find

$$g_i(\omega) = \frac{\theta}{\pi L} \int_0^\infty e^{-\tau/T} \cos \omega \tau d\tau = \frac{\theta}{\pi R} \frac{1}{1 + (\omega L/R)^2}. \quad (29.14)$$

Usually in metals $T = LR^{-1} \sim 10^{-13}$ sec, and one can write to a good approximation that

$$g_i(\omega) \approx \theta \pi^{-1} R^{-1}. \quad (29.15)$$

Thus, the mean-square value of the current in a closed circuit generated by a random e.m.f. in a frequency interval ω , $\omega + d\omega$ can be written in the form

$$\overline{i^2} d\omega = \theta \pi^{-1} R^{-1} d\omega. \quad (29.16)$$

In the case of an open circuit, instead of (29.16) one can write for the spontaneous fluctuations of e.m.f. the relation

$$\overline{\mathcal{E}^2} d\omega = \theta R \pi^{-1} d\omega. \quad (29.17)$$

Formulae (29.17) and (29.16) are called the Nyquist formulae. They allow one to take into account fluctuation phenomena in calculations on quasi-stationary circuits, for example, the fluctuation e.m.f., besides other macroscopic characteristics.

According to (29.16) the fluctuation current is proportional to $\theta^{\frac{1}{2}}$ and inversely proportional to $R^{\frac{1}{2}}$. This has a simple meaning. The resistance $R = \sigma^{-1} \sim n^{-1}$, where n is the number of electrons per cm^3 . In correspondence with the general formula (3.7) of Part III, $(\overline{i^2})^{\frac{1}{2}} \sim n^{\frac{1}{2}}$.

The above derivation was made under an essential restriction: the resistance has been calculated without taking into account the frequency. In Part VI we shall return to the Nyquist formula for the region of high frequencies. A detailed calculation will be made of the quantum effects at frequencies $\omega > kt/h \equiv \theta/h$.

In semiconductors an additional noise-producing mechanism arises which will also be discussed in Part VI.

§ 30. Skin effect

We have, up to now, studied quasistationary currents in linear circuits, considering the conductors to be infinitesimally thin. We shall now consider the distribution of alternating current over the cross-section of the conductor. Assuming, as before, that the conditions of quasistationarity are fulfilled, we write Maxwell's equations in a homogeneous conducting medium:

$$\nabla \times \mathbf{E} = -\frac{\mu}{c} \frac{\partial \mathbf{H}}{\partial t},$$

$$\nabla \times \mathbf{H} = \frac{4\pi\sigma}{c} \mathbf{E},$$

$$\nabla \cdot \mathbf{H} = 0,$$

$$\nabla \cdot \mathbf{E} = 0.$$

It is easy to obtain separate equations for the electric and magnetic fields. Taking the curl of curl \mathbf{E} , we find

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\nabla^2 \mathbf{E} = \frac{\mu}{c} \frac{\partial}{\partial t} (\nabla \times \mathbf{H}),$$

or

$$\nabla^2 \mathbf{E} = \frac{4\pi\sigma\mu}{c^2} \frac{\partial \mathbf{E}}{\partial t}. \quad (30.1)$$

A similar equation is obtained for the magnetic field.

Eq. (30.1) and the analogous equation for the vector \mathbf{H} determine the dependence of the fields on time and coordinates in the space occupied by the conductor. At the boundary of this space, i.e. at the surface of the conductor, the field vectors satisfy the usual boundary conditions.

We restrict ourselves to the solution of the field equations in the simple case of alternating current flowing in a conductor which occupies a half-space $z > 0$. The direction of the current is along the x -axis and its dependence on time is assumed to be defined:

$$j_x = j(z) e^{i\omega t}; \quad j_y = j_z = 0.$$

We note that j_x cannot depend on the coordinate x by virtue of the continuity equation which gives

$$\frac{\partial j_x}{\partial x} = 0.$$

In correspondence with Ohm's law we seek the electric field satisfying eq. (30.1) in the form

$$E_x = E(z) e^{i\omega t}; \quad E_y = E_z = 0. \quad (30.2)$$

The substitution of (30.2) into (30.1) gives

$$\frac{d^2 E(z)}{dz^2} = i \frac{4\pi\mu\sigma\omega}{c^2} E(z). \quad (30.3)$$

The general solution of the last equation is

$$E(z) = A e^{-kz} + B e^{kz},$$

where

$$k = \left(i \frac{4\pi\mu\sigma\omega}{c^2} \right)^{\frac{1}{2}} = \frac{1+i}{\sqrt{2}} \left(\frac{4\pi\mu\sigma\omega}{c^2} \right)^{\frac{1}{2}}.$$

We introduce the notation

$$\delta = \frac{c}{(2\pi\mu\sigma\omega)^{\frac{1}{2}}}. \quad (30.4)$$

Then

$$E(z) = A e^{-iz/\delta} e^{-z/\delta} + B e^{iz/\delta} e^{z/\delta}.$$

The value of B must, obviously, be equal to zero, in order that the field may have a finite value everywhere.

Thus, finally,

$$E_x = A e^{-z/\delta} e^{-i(z/\delta - \omega t)}. \quad (30.5)$$

Formula (30.5) shows that the electric field strength decreases exponentially inside a conductor.

An effective decrease in the field strength (decrease by a factor of e) takes place at a distance δ from the surface.

Knowing the electric field distribution in the conductor, one can find the magnetic field distribution.

We have, obviously,

$$-i \frac{\mu\omega}{c} H_y = (\nabla \times \mathbf{E})_y = \frac{\partial E_x}{\partial z} = -\frac{(i+1)}{\delta} E_x,$$

$$H_x = H_z = 0.$$

Hence

$$H_y = (1-i) \frac{c}{\mu\omega\delta} A e^{-i(z/\delta - \omega t)} e^{-z/\delta}. \quad (30.6)$$

The magnetic field turns out to be perpendicular to the electric field. It decreases inside the conductor according to the same law as the electric field. In absolute value $|H_y| \approx (c/\omega\delta) |E_x|$, i.e. is larger than the electric field by a factor λ/δ .

Thus, the electromagnetic field and correspondingly the entire current in a conductor turn out to be localized in a thin surface layer of thickness δ . The localization of the field in the thin surface layer is called the skin effect, and the quantity δ is called the skin depth.

It is easily seen that the entire Joule heat j^2/σ is released in the region of the skin depth.

From the definition of δ it is clear that its numerical value can vary within a wide range for different frequencies ω and conductivities σ . In order to get an idea of the order of magnitude of this quantity, we point out that for copper $\delta \approx 1$ cm for a frequency of 50 hertz and $\delta \approx 3 \times 10^{-3}$ cm for frequencies $\sim 10^5$ hertz.

As $\omega \rightarrow 0$, i.e. in the transition to direct current, $\delta \rightarrow \infty$ and the skin effect vanishes. The current is uniformly distributed over the entire cross-section of the conductor. On the other hand, in the transition to the limit $\sigma \rightarrow \infty$ the thickness of the skin layer tends to zero. This represents the case of the ideal conductor.

The result which we have obtained for the case of a simplified model of a conductor is of general character. For any geometric configuration of con-

auctors the field in them turns out to be localized in the skin depth.

If the distribution depends on the geometric properties of a conductor, then the problem of finding the field in it becomes only slightly more complicated than in the example discussed. Thus, in the case of a current flowing along a long cable of circular cross-section, solid or hollow, the current density vector is directed parallel to the generatrix of the cable. Hence the general trend of the solution of the problem of finding the field distribution in the cable is the same as in the example which we have discussed. For the skin depth one obtains the same numerical value. The geometric distribution of the field differs relatively little from an exponential decrease, particularly at high frequencies, when the radius of the cable is large in comparison with the skin depth. In the general case of a solid conductor of arbitrary form finding the alternating-field distribution represents a problem which is complex from the mathematical standpoint *. However, irrespective of the form of the conductor and even of the mechanism of excitation of the field in it (by means of an impressed e.m.f., an external alternating magnetic field etc.) the general inference remains valid: the alternating electromagnetic field penetrates into the conductor to the depth of the skin layer δ . In this case the magnetic field is larger than the electric field in the ratio λ/δ .

Analogous results can be obtained in the consideration of another problem. Let the metal be acted upon by an alternating electromagnetic field depending on time according to the law $e^{i\omega t}$.

If the magnetic field at the surface of the conductor $z = 0$ has a value H_0 , the boundary condition (5.4) allows one to formulate the boundary value problem

$$\frac{\partial^2 H(z, t)}{\partial z^2} = \frac{4\pi\sigma\mu}{c^2} \frac{\partial H(z, t)}{\partial t},$$

$$H = H_0 \quad \text{for} \quad z = 0,$$

$$H \rightarrow 0 \quad \text{for} \quad z \rightarrow \infty.$$

The solution of the boundary value problem can be written, analogously to (30.5), in the form

$$H(z, t) = H_0 e^{-z/\delta} e^{i\omega t},$$

where the thickness of the skin layer δ is given by formula (30.4).

* See W.R.Smythe, *Static and dynamic electricity* (McGraw-Hill, New York, 1950).

It should be emphasized that the skin effect becomes more pronounced in going to high frequencies. We shall see in the next chapter that in high-frequency fields, when the fields can no longer be considered as quasistationary, there is as before a skin effect, although the penetration depth turns out, as a rule, to be different (see §33). The skin effect plays an important role in alternating-current engineering. It allows one to use hollow cables or cables covered with a layer of metal of a particularly high conductivity, which reduces the expenditure of material and power.

High-frequency Fields

§31. Electromagnetic waves in a homogeneous isotropic medium

We consider the propagation of the electromagnetic field in a spatially homogeneous and isotropic medium characterized by material constants ϵ_0 , μ_0 and σ_0 , i.e. in a medium without spatial dispersion. The subscript zero denotes that the material constants have a static value and are referred to the frequency $\omega = 0$. We shall discuss below the conditions of applicability of this assumption.

We assume the medium to be non-ferromagnetic ($\mu_0 \approx 1$). We have seen above that for metals the penetration of the field is very small, even at relatively low frequencies. Therefore it makes no sense to consider the propagation of the electromagnetic field inside the metal. However, such a consideration is not out of place for media having conductivities which are lower than those for metals (e.g. semi-conductors or solutions of electrolytes). In the limiting case $\sigma \rightarrow 0$ we pass over to the case of ideal dielectrics. We can also assume that the magnetic permeability of the medium is equal to unity. Maxwell's equations then have the form

$$\nabla \times \mathbf{H} = \frac{4\pi\sigma}{c} \mathbf{E} + \frac{\epsilon_0}{c} \frac{\partial \mathbf{E}}{\partial t}, \quad (31.1)$$

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t}, \quad (31.2)$$

$$\nabla \cdot \mathbf{E} = 0, \quad (31.3)$$

$$\nabla \cdot \mathbf{H} = 0. \quad (31.4)$$

Taking the curl of the first equation and making use of the last, we have

$$\nabla \times (\nabla \times \mathbf{H}) = -\nabla^2 \mathbf{H} = \frac{4\pi\sigma}{c} (\nabla \times \mathbf{E}) + \frac{\epsilon_0}{c} \frac{\partial}{\partial t} (\nabla \times \mathbf{E}),$$

or, using (31.2),

$$\nabla^2 \mathbf{H} - \frac{\epsilon_0}{c^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} = \frac{4\pi\sigma}{c^2} \frac{\partial \mathbf{H}}{\partial t}. \quad (31.5)$$

Proceeding analogously with (31.2), we easily arrive at the equation

$$\nabla^2 \mathbf{E} - \frac{\epsilon_0}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = \frac{4\pi\sigma}{c^2} \frac{\partial \mathbf{E}}{\partial t}. \quad (31.6)$$

We seek solutions of eqs. (31.5) and (31.6) in the form of monochromatic plane waves propagating along the x -axis.

Consequently, we assume that the solutions of (31.5) and (31.6) have the form

$$\mathbf{H}(x, t) = \mathbf{H}_0(x) e^{i\omega t}, \quad (31.7)$$

$$\mathbf{E}(x, t) = \mathbf{E}_0(x) e^{i\omega t}. \quad (31.8)$$

Substituting (31.7) into (31.5), we find

$$\frac{d^2 \mathbf{H}_0(x)}{dx^2} + k^2 \mathbf{H}_0(x) = 0, \quad (31.9)$$

where k denotes the complex quantity

$$k = \frac{\omega}{c} \left(\epsilon_0 - i \frac{4\pi\sigma}{\omega} \right)^{\frac{1}{2}}. \quad (31.10)$$

We introduce the important notion of the complex dielectric permittivity $\epsilon(\omega)$ in a conducting medium, defining it by the relation

$$k = \frac{\omega}{c} [\epsilon(\omega)]^{\frac{1}{2}} = k_v [\epsilon(\omega)]^{\frac{1}{2}}, \quad (31.11)$$

where $k_v = \omega/c$ is the wave number in vacuum.

From (31.10) it follows that

$$\epsilon(\omega) = \epsilon_0 - i \frac{4\pi\sigma}{\omega}. \quad (31.12)$$

This important formula establishes the relation between the dielectric permittivity and the conductivity. Further, we write the complex quantity $[\epsilon(\omega)]^{\frac{1}{2}}$ in the form

$$[\epsilon(\omega)]^{\frac{1}{2}} = n - i\kappa. \quad (31.13)$$

The quantities n and κ are called, respectively, the refractive index and the absorption coefficient. The sign of κ is chosen in such a way that the imaginary part of k is essentially negative (for $\kappa > 0$). The reason for such a choice is obvious from what follows.

The solutions (31.7) have the form

$$\mathbf{H} = \mathbf{A}_1 e^{-k_v \kappa x} e^{i(\omega t - k_v n x)}, \quad (31.14)$$

where the vector \mathbf{A}_1 is the complex amplitude. Analogously, for the electric field one can write

$$\mathbf{E} = \mathbf{A}_2 e^{-k_v \kappa x} e^{i(\omega t - k_v n x)}. \quad (31.15)$$

In the case of arbitrary direction of propagation the field vectors can be written in the form

$$\mathbf{H} = \mathbf{A}_1 e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}, \quad (31.16)$$

$$\mathbf{E} = \mathbf{A}_2 e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}, \quad (31.17)$$

where \mathbf{k} is the wave vector, $\mathbf{k} = k\mathbf{k}_0$, and \mathbf{k}_0 is a unit vector in the direction of propagation of the wave. The relation between the vectors \mathbf{H} and \mathbf{E} can be

$$\nabla \times \mathbf{E} = -i(\mathbf{k} \times \mathbf{E}) = -ik(\mathbf{k}_0 \times \mathbf{E}) = -i \frac{\omega}{c} [\epsilon(\omega)]^{\frac{1}{2}} (\mathbf{k}_0 \times \mathbf{E}) = -i \frac{\omega}{c} \mathbf{H}.$$

Hence we find

$$\mathbf{H} = [\epsilon(\omega)]^{\frac{1}{2}} (\mathbf{k}_0 \times \mathbf{E}). \quad (31.18)$$

In contrast to the case of the propagation of electromagnetic waves in vacuum, the amplitudes of the electric and magnetic field here turn out to be different. However, as in a vacuum, electromagnetic waves in a homogeneous and isotropic medium are transverse. Indeed, substituting (31.16) and (31.17) into (31.3) and (31.4), we find

$$\mathbf{k} \cdot \mathbf{E} = \mathbf{k} \cdot \mathbf{H} = 0. \quad (31.19)$$

Formulae (31.16) – (31.19) show that Maxwell's equations allow solutions in the form of transverse plane waves with wave number k and an arbitrary frequency ω in a homogeneous and isotropic medium, as in vacuum. However damping of the wave according to an exponential law occurs in the medium. The effectiveness of the damping is determined by the quantity κ . We find the values of n and κ from (31.12) and (31.13). Squaring (31.13), substituting in (31.12) and separating the real and imaginary parts, we obtain

$$n^2 - \kappa^2 = \epsilon_0, \quad \kappa n = 2\pi\sigma/\omega. \quad (31.20)$$

The solution of (31.20) gives

$$n = [\frac{1}{2}(\epsilon_0^2 + 16\pi^2\sigma^2/\omega^2)^{\frac{1}{2}} + \epsilon_0]^{\frac{1}{2}}, \quad (31.21)$$

$$\kappa = [\frac{1}{2}(\epsilon_0^2 + 16\pi^2\sigma^2/\omega^2)^{\frac{1}{2}} - \epsilon_0]^{\frac{1}{2}}. \quad (31.22)$$

The signs of the roots are chosen in such a way that n and κ have real values and, in addition, that κ is positive.

Formulae (31.21) and (31.22) determine the law of dispersion in a conducting medium. It should be noted that at sufficiently high frequencies the conductivity σ is also dependent on the frequency.

Let us consider the limiting cases of formulae (31.21) and (31.22). If the inequality

$$\sigma \ll \epsilon_0 \omega / 4\pi \quad (31.23)$$

is valid, then this means that the conduction current $\sigma \mathbf{E}$ is small in comparison with the displacement current $(\epsilon_0/4\pi)(\partial \mathbf{E}/\partial t) \sim (\epsilon_0 \omega/4\pi) \mathbf{E}$. This holds for ideal dielectrics (for which $\sigma \rightarrow 0$) as well as for real dielectrics possessing a very low conductivity, and also for conductors of non-metallic type (semiconductors, electrolytes) at sufficiently high frequencies.

In this case one can simplify formulae (31.21) and (31.22), writing them in the form

$$n \approx \epsilon_0^{\frac{1}{2}}, \quad (31.24)$$

$$\kappa \approx 2\pi\sigma/\omega\epsilon_0^{\frac{1}{2}}. \quad (31.25)$$

By virtue of the inequality (31.23), the following inequality also holds:

$$n \gg \kappa.$$

If one can completely disregard the quantity κ which, strictly speaking, is equal to zero only in an ideal dielectric, then the medium is called transparent. For a transparent medium the formulae for the field vectors are simplified and assume the form

$$\mathbf{E} = \mathbf{A} \exp \left[i \left(\omega t - \frac{\omega}{c} \epsilon_0^{\frac{1}{2}} \mathbf{k}_0 \cdot \mathbf{r} \right) \right] = \mathbf{A} \exp [i(\omega t - \mathbf{k} \cdot \mathbf{r})], \quad (31.26)$$

$$\mathbf{H} = \epsilon_0^{\frac{1}{2}} \mathbf{A} \exp \left[i \left(\omega t - \frac{\omega}{c} \epsilon_0^{\frac{1}{2}} \mathbf{k}_0 \cdot \mathbf{r} \right) \right] = \epsilon_0^{\frac{1}{2}} \mathbf{A} \exp [i(\omega t - \mathbf{k} \cdot \mathbf{r})], \quad (31.27)$$

$$\mathbf{k} = \frac{\omega}{c} \epsilon_0^{\frac{1}{2}} \mathbf{k}_0 = \frac{\omega}{v} \mathbf{k}_0, \quad (31.28)$$

where v is the phase velocity of propagation of the waves

$$v = c\epsilon_0^{-\frac{1}{2}}. \quad (31.29)$$

The last formula justifies the name of the quantity n . Indeed, as is well known (see §37), the refractive index means the ratio of the velocity of propagation of electromagnetic waves in vacuum to that in the medium:

$$n = \frac{c}{v} = \epsilon_0^{\frac{1}{2}}. \quad (31.30)$$

Electromagnetic waves in a non-conducting medium differ from those in vacuum only in the velocity of propagation, which is smaller by a factor of $\epsilon_0^{\frac{1}{2}}$ than the velocity of light in vacuum c . Moreover, the amplitudes of the electric and magnetic fields are in the ratio

$$\frac{|H|}{|E|} = \epsilon_0^{\frac{1}{2}} = n. \quad (31.31)$$

The above ratios, established by Maxwell, played an important role in the history of the development of the theory of the electromagnetic field. In particular, formula (31.30) established a relation between electromagnetic and optical phenomena, which before the work of Faraday and Maxwell were considered to be completely independent.

Formula (31.30) was subjected to extensive experimental test for a large number of liquids and gases. Good agreement with experiment was found for a number of liquids and gases in the visible and infrared parts of the spectrum. However, the relation (31.30) is not satisfied at all for liquids (e.g. water) and gases whose molecules have a considerable intrinsic dipole moment. Other restrictions on the applicability of (31.30) are associated with the existence of absorption and dispersion.

In the case of weak absorption it is convenient to write $\epsilon^{\frac{1}{2}}$ in the form

$$[\epsilon(\omega)]^{\frac{1}{2}} = (n^2 + \kappa^2)^{\frac{1}{2}} \exp [-i \arctan (\kappa/n)] \quad (31.32)$$

and to write the formulae for the field vectors in the form

$$\mathbf{E} = \mathbf{A} \exp [-\kappa \mathbf{k}_0 \cdot \mathbf{r}] \exp [i(\omega t - n \mathbf{k}_0 \cdot \mathbf{r})], \quad (31.33)$$

$$\mathbf{H} \approx n(\mathbf{k}_0 \times \mathbf{E}) \exp [-i \arctan (\kappa/n)]. \quad (31.34)$$

The magnetic field strength differs in phase from the electric field strength by an amount $\arctan (\kappa/n)$.

In addition we shall find the time average of the electromagnetic field energy density in a weakly absorbing medium. We have, obviously,

$$\overline{\frac{\epsilon_0}{8\pi} E^2} = \frac{\epsilon_0 A^2}{16\pi} e^{-2\kappa x},$$

$$\overline{\frac{H^2}{8\pi}} = \frac{n^2 + \kappa^2}{16\pi} A^2 e^{-2\kappa x} \approx \frac{\epsilon_0 A^2}{16\pi} e^{-2\kappa x},$$

since $n^2 + \kappa^2 \approx n^2 = \epsilon_0$. As in vacuum, the energies of the electric and magnetic fields in the wave are nearly equal to each other.

Let us now consider the opposite limiting case where the conduction current is large in comparison with the displacement current. Formally, this case is obtained immediately from the dispersion formulae for low frequencies and large conductivity. We have, obviously,

$$\kappa \approx (2\pi\sigma/\omega)^{\frac{1}{2}}, \quad (31.35)$$

$$n \approx \kappa \approx (2\pi\sigma/\omega)^{\frac{1}{2}}. \quad (31.36)$$

As is shown by comparison of (31.35) with (30.4), the attenuation of the field in a medium takes place over the thickness of its skin depth. This represents the case of strong absorption.

It is clear, however, that these formulae have a somewhat relative character, since it is difficult to speak of the propagation of waves if they are damped over the thickness of the skin layer.

In conclusion, let us write the energy conservation law for a medium with frequency dispersion. Assuming the medium to be in a state of thermodynamic equilibrium at constant temperature, we can write for the change of internal energy per unit volume

$$\frac{dU_0}{dt} = \frac{dQ}{dt} + \frac{dW}{dt} = \frac{dQ}{dt} + \frac{\mathbf{E}}{8\pi} \frac{d\mathbf{D}}{dt}. \quad (31.37)$$

We average eq. (31.37) over time, assuming that the temperature is maintained constant. Then, on the average, the energy of the medium in an external periodic field remains constant, so that for the heat released we obtain

$$\overline{\frac{dQ}{dt}} = -\overline{\frac{\mathbf{E}}{8\pi} \frac{d\mathbf{D}}{dt}}.$$

Setting

$$\mathbf{D} = [\epsilon_1(\omega) + i\epsilon_2(\omega)] (\mathbf{E}_0 \cos \omega t + i\mathbf{E}_0 \sin \omega t),$$

we have

$$\frac{dQ}{dt} = \frac{\epsilon_2 \omega E_0^2}{8\pi}. \quad (31.38)$$

Thus release of heat determined by the quantity $\epsilon_2(\omega)$ occurs in the medium. Formula (31.38) makes sense if the heat released can be removed and the temperature of the medium maintained constant. Since heat is emitted only, Q is always non-negative and, correspondingly,

$$\epsilon_2 \geq 0. \quad (31.39)$$

The sign of ϵ_1 can be either positive or negative.

§32. Dispersion relations (the Kramers—Kronig formulae)

We have already pointed out in §6 that the phenomenon of dispersion occurs at high frequencies, when the frequency of the radiation is comparable with characteristic atomic frequencies and its wavelength is comparable with the dimensions of spatial non-uniformities in the medium.

In the case of a homogeneous and isotropic medium the general formula (6.5) can be written in the form

$$\mathbf{D}(\mathbf{r}, t) = \int_{-\infty}^t dt' \int \epsilon(|\mathbf{r}-\mathbf{r}'|, t-t') \mathbf{E}(\mathbf{r}', t') dV'. \quad (32.1)$$

The function $\epsilon(|\mathbf{r}-\mathbf{r}'|, t-t')$ depends only on the difference between the coordinates, since in a homogeneous medium there are no unique points. A contribution to \mathbf{D} is given by the change in polarization which is transferred in a time $t-t'$ over a distance $|\mathbf{r}-\mathbf{r}'|$. The integral is taken only with respect to the past, in order to satisfy the principle of causality. It is obvious that $|\mathbf{r}-\mathbf{r}'| < c(t-t')$, but for simplicity we shall not take into account this restriction*.

We expand the electric field and induction in Fourier integrals:

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \int \mathbf{E}(\mathbf{k}, \omega) e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} d\mathbf{k} d\omega, \quad (32.2)$$

$$\mathbf{D}(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \int \mathbf{D}(\mathbf{k}, \omega) e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} d\mathbf{k} d\omega. \quad (32.3)$$

Then, substituting into (32.1), we find

$$\mathbf{D}(\mathbf{k}, \omega) = \epsilon(\mathbf{k}, \omega) \mathbf{E}(\mathbf{k}, \omega), \quad (32.4)$$

* See M.A. Leontovich, Soviet Physics JETP 13 (1961) 634.

where

$$\begin{aligned}\epsilon(\mathbf{k}, \omega) &= \int_{-\infty}^t dt' \int \epsilon(|\mathbf{r}-\mathbf{r}'|, t-t') e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} e^{-i\omega(t-t')} dV' = \\ &= \int_{-\infty}^{\infty} d\tau \int \epsilon(\mathbf{R}, \tau) e^{-i(\mathbf{k}\cdot\mathbf{R}-\omega\tau)} d\mathbf{R}.\end{aligned}\quad (32.5)$$

Here new variables $\mathbf{R} = \mathbf{r} - \mathbf{r}'$ and $\tau = t - t'$ are introduced. The function $\epsilon(\mathbf{k}, \omega)$ is not the Fourier transform of $\epsilon(\mathbf{R}, \tau)$, since the integral with respect to τ is taken in the range from zero to infinity. It is the function $\epsilon(\mathbf{k}, \omega)$ (and not $\epsilon(\mathbf{R}, \tau)$) which is of basic importance, because, according to (32.4), it relates the vectors \mathbf{D} and \mathbf{E} . It is called the dielectric permittivity, and depends on the frequency and the wave vector.

We note that the dependence of $\epsilon(\mathbf{k}, \omega)$ on the wave vector \mathbf{k} is associated with the dependence of $\epsilon(\mathbf{R}, \tau)$ on the space coordinates \mathbf{R} , i.e. with the taking into account of spatial dispersion.

If the wavelength is small compared with the spatial non-uniformities, then one can assume the dielectric constant to be a function of $t - t'$ only and, instead of (32.5), one can write the simpler relation

$$\epsilon(\omega) = \int_0^{\infty} \epsilon(\tau) e^{-i\omega\tau} d\tau, \quad (32.6)$$

in which the dielectric constant depends only on the frequency. In this case the relation (32.4) is simplified and becomes

$$\mathbf{D}(\omega) = \epsilon(\omega) \mathbf{E}(\omega). \quad (32.7)$$

We note that if the dielectric constant depends on the frequency, then only eq. (32.7) represents the correct form of notation of the constitutive equation. The relation

$$\mathbf{D}(\mathbf{r}, t) = \epsilon(\omega) \mathbf{E}(\mathbf{r}, t)$$

(where \mathbf{D} and \mathbf{E} are the vectors themselves and not their Fourier amplitudes, as in (32.4)) which is often presented in old text-books, makes no sense. If ϵ is a function of the frequency, then eq. (32.4) must contain quantities referring to this frequency.

The function $\epsilon(\mathbf{k}, \omega)$ is, in general, complex and can be written in the form

$$\epsilon(\mathbf{k}, \omega) = \epsilon^{\text{re}}(\mathbf{k}, \omega) + i\epsilon^{\text{im}}(\mathbf{k}, \omega).$$

It is obvious that the real and imaginary parts are

$$\epsilon^{\text{re}}(\mathbf{k}, \omega) = \frac{\epsilon(\mathbf{k}, \omega) + \epsilon^*(\mathbf{k}, \omega)}{2}, \quad (32.8)$$

$$\epsilon^{\text{im}}(\mathbf{k}, \omega) = \frac{\epsilon(\mathbf{k}, \omega) - \epsilon^*(\mathbf{k}, \omega)}{2i}.$$

Moreover, from the definition (32.5) it follows that

$$\epsilon(\mathbf{k}, \omega) = \epsilon^*(-\mathbf{k}, -\omega). \quad (32.9)$$

Substituting the expression (32.5) into (32.8) and taking into account (32.9), we find

$$\epsilon^{\text{re}}(\mathbf{k}, \omega) = \frac{1}{2} \int_{-\infty}^{\infty} d\tau \int \epsilon(|\mathbf{R}|, |\tau|) e^{-i(\mathbf{k} \cdot \mathbf{R} - \omega\tau)} d\mathbf{R}. \quad (32.10)$$

Analogously,

$$\begin{aligned} \epsilon^{\text{im}}(\mathbf{k}, \omega) &= \frac{1}{2i} \int_0^{\infty} d\tau \int \epsilon(|\mathbf{R}|, |\tau|) e^{-i(\mathbf{k} \cdot \mathbf{R} - \omega\tau)} d\mathbf{R} \\ &\quad - \frac{1}{2i} \int_{-\infty}^0 d\tau \int \epsilon(|\mathbf{R}|, |\tau|) e^{-i(\mathbf{k} \cdot \mathbf{R} - \omega\tau)} d\mathbf{R} = \\ &= \frac{1}{2i} \int_{-\infty}^{\infty} d\tau \int \epsilon(|\mathbf{R}|, |\tau|) e^{-i(\mathbf{k} \cdot \mathbf{R} - \omega\tau)} \text{sgn } \tau d\mathbf{R}, \end{aligned} \quad (32.11)$$

where the sign function $\text{sgn } \tau$ is determined according to formula (III.16). Since the real and imaginary parts of $\epsilon(\mathbf{k}, \omega)$ are expressed in terms of $\epsilon(|\mathbf{R}|, |\tau|)$ according to formulae (32.10) and (32.11), they can be interrelated.

That is, since formula (32.11) represents the Fourier transform of $(\text{sgn } \tau \cdot \epsilon)$, then taking the inverse of this transform we have

$$\text{sgn } \tau \cdot \epsilon(|\mathbf{R}|, |\tau|) = \frac{i}{\pi} \int \epsilon^{\text{im}}(\mathbf{k}', \omega') e^{i(\mathbf{k}' \cdot \mathbf{R} - \omega'\tau)} d\mathbf{k}' d\omega'. \quad (32.12)$$

Substituting the value of $\epsilon(|\mathbf{R}|, |\tau|)$ from (32.12) into (32.10), we find

$$\begin{aligned}
 \epsilon^{\text{re}}(\mathbf{k}, \omega) &= \frac{i}{(2\pi)^4} \int_{-\infty}^{\infty} d\tau \int d\omega' \int d\mathbf{k}' \int d\mathbf{R} \epsilon^{\text{im}}(\mathbf{k}', \omega') \operatorname{sgn} \tau e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{R}} e^{-i(\omega - \omega')\tau} \\
 &= \frac{i}{(2\pi)^4} \int d\omega' \int d\mathbf{k}' \epsilon^{\text{im}}(\mathbf{k}', \omega') \int_{-\infty}^{\infty} e^{-i(\omega - \omega')\tau} \operatorname{sgn} \tau d\tau \int e^{i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{R}} d\mathbf{R} \\
 &= \frac{i}{2\pi} \int d\omega' \int d\tau \operatorname{sgn} \tau e^{-i(\omega - \omega')\tau} \int \epsilon^{\text{im}}(\mathbf{k}', \omega') \delta(\mathbf{k} - \mathbf{k}') d\mathbf{R} = \\
 &= \frac{i}{2\pi} \int d\omega' \epsilon^{\text{im}}(\mathbf{k}, \omega') \int_{-\infty}^{\infty} e^{-i(\omega - \omega')\tau} \operatorname{sgn} \tau d\tau. \quad (32.13)
 \end{aligned}$$

The inner integral can be calculated by making use of formula (III.20) for the sign function (cf. Volume I, Appendix III)

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{sgn} \tau e^{i(\omega - \omega')\tau} d\tau = \frac{i}{\pi} \mathcal{P} \left(\frac{1}{\omega - \omega'} \right).$$

Hence

$$\epsilon^{\text{re}}(\mathbf{k}, \omega) = \frac{1}{\pi} \mathcal{P} \int \frac{\epsilon^{\text{im}}(\mathbf{k}, \omega')}{\omega - \omega'} d\omega'. \quad (32.14)$$

The integral in formula (32.14) is taken in the sense of a principal value. Analogous calculations – the inversion of formula (32.10) and the substitution into (32.11) – give

$$\epsilon^{\text{im}}(\mathbf{k}, \omega) = -\frac{1}{\pi} \mathcal{P} \int \frac{\epsilon^{\text{re}}(\mathbf{k}, \omega')}{\omega - \omega'} d\omega'. \quad (32.15)$$

Formulae (32.14) and (32.15) are called the Kramers–Kronig formulae or dispersion relations. They relate the real and imaginary parts of the dielectric constant, in other words, the characteristics of the processes of dispersion and absorption. All the above calculations can be applied to the constitutive equation of $\mu(\mathbf{k}, \omega)$, and the same relations can be written for the real and imaginary parts of the magnetic permeability. Thus, the general principle according to which the real and imaginary parts of the basic quantities which characterize the properties of a medium – the permittivity $\epsilon(\mathbf{k}, \omega)$ and the

permeability $\mu(\mathbf{k}, \omega)$ — are correlated is established. Since, according to (31.12), the electrical conductivity is connected with the dielectric constant, the result obtained holds equally for electrical conductivity.

The Kramers—Kronig formulae are among the most general relations of electrodynamics. Indeed, in deriving them only two assumptions have been made:

- (1) the existence of the causal relationship: the value of the function \mathbf{D} at an instant t can be determined by polarization processes which occurred only at preceding instants of time,
- (2) the possibility of the expansion of all the functions in a Fourier integral.

The latter assumption is in practice fulfilled for all functions describing physical processes. Thus, the Kramers—Kronig formulae establish the relation between quantities which characterize dispersion processes (the real part of ϵ , μ or σ) and absorption processes (the imaginary part of these quantities) in the most general form. Dispersion processes and absorption processes turn out to be interrelated. Illustrations of this general proposition are encountered in many different branches of physics, in particular in quantum mechanics and the theory of elementary particles (see Part V). Besides the importance of this principle, the Kramers—Kronig formulae are of great practical importance. The imaginary part of the dielectric constant is connected with the absorption of energy in the medium (see (31.13)) and can be found relatively simply by experiment, while the real part ϵ^{re} is found from the Kramers—Kronig formula.

§33. The electromagnetic field in a medium with spatial and time dispersion

The solution of Maxwell's equations which we have already obtained in the form of transverse electromagnetic waves turns out to be not only qualitatively but also quantitatively inapplicable in the case of the propagation of electromagnetic waves in a medium with spatial dispersion.

As we have emphasized in §6, the phenomenon of spatial dispersion proves to be important in the case where the wavelength becomes comparable with the size of the spatial non-uniformities of the medium. Then the value of the induction at a given point \mathbf{r} of space turns out, according to (6.5), to be dependent on the value of the field vector and the properties of the medium in the region of space surrounding it (the whole set of points \mathbf{r}').

In this case one usually speaks of the "non-local connection" between the corresponding quantities \mathbf{D} and \mathbf{E} .

Until recently it was assumed that in macroscopically homogeneous media

the phenomenon of spatial dispersion could be disregarded. However, it turns out that spatial dispersion plays an important role in a very large class of macroscopically homogeneous media called plasma-like media. This class comprises the plasma itself (whose properties will be discussed in ch. 6) as well as conductors (metals) and semi-conductors with a high electric conductivity, when one has to investigate their interaction with high-frequency fields.

In such media there are characteristic non-uniformities of a relatively large scale. As an example one can mention the mean free path of electrons in metals, which will be calculated in Part VI. The mean free path amounts to about 7×10^{-6} cm – 10^{-5} cm.

Another example of spatial non-uniformity in connection with a plasma will be presented in §46.

The important part played by plasma-like media in contemporary physics has made the study of the properties of media with spatial dispersion very relevant.

Let us consider a macroscopically homogeneous, and isotropic, non-magnetizable ($\mu \approx 1$) medium without currents or free charges. We seek a solution of Maxwell's equations in the form of expansions in a Fourier integral. This method has now become one of the most popular and frequently used methods. We write the expansions in the form

$$\mathbf{D}(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \int e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \mathbf{D}(\mathbf{k}, \omega) d\mathbf{k} d\omega, \quad (33.1)$$

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \int e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \mathbf{E}(\mathbf{k}, \omega) d\mathbf{k} d\omega, \quad (33.2)$$

$$\mathbf{B}(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \int e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \mathbf{B}(\mathbf{k}, \omega) d\mathbf{k} d\omega, \quad (33.3)$$

$$\mathbf{H}(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \int e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \mathbf{H}(\mathbf{k}, \omega) d\mathbf{k} d\omega. \quad (33.4)$$

Substituting these expansions into Maxwell's equations, we find

$$i [\mathbf{k} \times \mathbf{H}(\mathbf{k}, \omega)] = -i \frac{\omega}{c} \mathbf{D}(\mathbf{k}, \omega), \quad (33.5)$$

$$i [\mathbf{k} \times \mathbf{E}(\mathbf{k}, \omega)] = i \frac{\omega}{c} \mathbf{B}(\mathbf{k}, \omega), \quad (33.6)$$

$$\mathbf{k} \cdot \mathbf{D} = 0, \quad (33.7)$$

$$\mathbf{k} \cdot \mathbf{B} = 0. \quad (33.8)$$

Thus the problem of integration of the system of equations in partial derivatives is reduced to the problem of the solution of a system of algebraic equations and the subsequent inversion of the Fourier transformation formulae.

In order that these relations assume a practical meaning, one has to write constitutive equations for $\mathbf{D}(\mathbf{k}, \omega)$ and $\mathbf{E}(\mathbf{k}, \omega)$, and $\mathbf{B}(\mathbf{k}, \omega)$ and $\mathbf{H}(\mathbf{k}, \omega)$ respectively.

We introduce the dielectric constant connecting $\mathbf{D}(\mathbf{k}, \omega)$ and $\mathbf{E}(\mathbf{k}, \omega)$. We note that, since the dielectric constant depends on \mathbf{k} , even for an isotropic medium it is a tensor and not a scalar.

That is, since it can depend on the vector \mathbf{k} , and the direction of the vector \mathbf{k} is the only direction defined in the isotropic medium, in order to determine the dielectric constant we have to form a symmetric tensor of the second rank containing only the tensor δ_{ij} and the components of the vector \mathbf{k} .

The only second-rank symmetric tensor ϵ_{ij} which can be made up of these quantities has the form

$$\epsilon_{ij}(\mathbf{k}, \omega) = \epsilon_{\perp}(\mathbf{k}, \omega) \left(\delta_{ij} - \frac{k_i k_j}{k^2} \right) + \epsilon_{\parallel} \frac{k_i k_j}{k^2}. \quad (33.9)$$

If the direction of the wave vector is chosen as the z -axis, then one can write for ϵ_{ij} the explicit expression

$$\epsilon_{ij} = \begin{vmatrix} \epsilon_{\perp} & 0 & 0 \\ 0 & \epsilon_{\perp} & 0 \\ 0 & 0 & \epsilon_{\parallel} \end{vmatrix}. \quad (33.10)$$

An isotropic medium with spatial dispersion is characterized by two dielectric constants: the longitudinal one ϵ_{\parallel} and the perpendicular one ϵ_{\perp} . The reason for such a terminology will be clear from what follows. It goes without saying that ϵ_{\parallel} and ϵ_{\perp} are functions of \mathbf{k} and ω .

If spatial dispersion is absent, then the dielectric constant $\epsilon(t-t')$ depends only on the time. Correspondingly, ϵ_{ij} in (33.9) must depend only on ω and not on \mathbf{k} . It can then be assumed that

$$\epsilon_{\perp}(\omega) = \epsilon_{\parallel}(\omega) = \epsilon(\omega). \quad (33.11)$$

Then

$$\epsilon_{ij} = \epsilon(\omega) \delta_{ij}. \quad (33.12)$$

The constitutive equation for the Fourier components will have the form

$$D_i(\mathbf{k}, \omega) = \epsilon_{ij}(\mathbf{k}, \omega) E_j(\mathbf{k}, \omega). \quad (33.13)$$

Forming the vector product of (33.6) and \mathbf{k} , we have

$$\mathbf{k} \times (\mathbf{k} \times \mathbf{E}) = k^2 \mathbf{E} - \mathbf{k}(\mathbf{k} \cdot \mathbf{E}) = -\frac{\omega}{c} (\mathbf{k} \times \mathbf{B}).$$

Substituting into the right-hand side its value from (33.5), we find

$$k^2 \mathbf{E} - \mathbf{k}(\mathbf{k} \cdot \mathbf{E}) = \frac{\omega^2}{c^2} \mathbf{D}, \quad (33.14)$$

or

$$k^2 E_j \delta_{ij} - k_i (k_j E_j) = \frac{\omega^2}{c^2} D_i.$$

Making use of the constitutive equation (33.13), we finally obtain

$$\left\{ k^2 \delta_{ij} - k_i k_j - \frac{\omega^2}{c^2} \epsilon_{ij} \right\} E_j = 0. \quad (33.15)$$

Formula (33.15) determines a system of linear algebraic equations. In order that this system should have a non-trivial solution it is necessary that the following condition be fulfilled:

$$k^2 \delta_{ij} - k_i k_j - \frac{\omega^2}{c^2} \epsilon_{ij} = 0. \quad (33.16)$$

This expression determines the dependence of ϵ_{ij} on \mathbf{k} and ω , i.e. represents a dispersion equation. We note, first of all, that taking into account the definition (33.9) and choosing the direction of the vector \mathbf{k} as the z -axis ($k_x = k_y = 0$, $k_z = k$), one can write the equation in the form of a system of algebraic equations

$$\frac{\omega^2}{c^2} \epsilon_{\perp}(\mathbf{k}, \omega) - k^2 = 0, \quad (33.17)$$

$$\epsilon_{\parallel} = \epsilon_{zz} = 0. \quad (33.18)$$

Eqs. (33.18) and (33.17) are independent dispersion equations for electromagnetic waves which can propagate in a medium with spatial dispersion. We see that in such a medium the existence of two independent wave processes is possible: transverse waves, for which the dispersion law (33.17) is valid, and longitudinal waves, for which the dispersion law is given by formula (33.18). For transverse waves the electric field vector \mathbf{E}_{\perp} has components E_x and E_y which are different from zero; for longitudinal waves the field has only one component: E_z . In order to avoid misunderstanding, we point out that for longitudinal waves the condition (33.7) is fulfilled, so that the vector \mathbf{D} is perpendicular to the vector \mathbf{k} , i.e. to the direction of propagation. However, the vectors \mathbf{D} and \mathbf{E} , by virtue of (33.14), are not parallel to each other. The appearance of longitudinal electromagnetic waves, which are often called polarization waves, appears to be a specific effect associated with the spatial dispersion of a medium. Longitudinal electromagnetic waves in a medium with spatial dispersion have a simple meaning. Consider a medium with an equilibrium but non-uniform charge distribution and assume that such a medium is acted upon by an electromagnetic field with a wavelength which is comparable to the size of the non-uniformity. The field causes a displacement of the charges (i.e. produces a polarization), violating the equilibrium distribution. As a result, charge oscillations which are very similar to elastic sound waves in isotropic media arise in the medium.

A more concrete quantitative description of the propagation of longitudinal waves will be developed in §46 for the case of a plasma. There we shall obtain explicit expressions for ϵ_{\perp} and ϵ_{\parallel} and, by means of eqs. (33.18) and (33.17), the law of dispersion in a plasma, also obtain $\omega_{\perp}(\mathbf{k})$ and $\omega_{\parallel}(\mathbf{k})$ for transverse and longitudinal waves.

In the case of a spatially homogeneous medium the dispersion equation goes over into the equation

$$\frac{\omega^2}{c^2} \epsilon(\omega) - k^2 = 0, \quad (33.19)$$

which is the same as (31.11) and which determines unambiguously the wave frequency $\omega(\mathbf{k})$.

Since, by virtue of (33.11), in such a medium $\epsilon_{||} = \epsilon(\omega) \neq 0$, the only possible solution of eq. (31.6) reads

$$E_{||} \equiv E_z = 0.$$

Thus, in correspondence with the results of §31, only transverse waves can propagate in a medium without spatial dispersion.

We note that we have obtained the dispersion equations and other basic properties of waves without finding the field vectors as functions of the coordinates and time. The latter require the inversion of the Fourier integrals, which, according to (33.1) – (33.4) and (33.13), can actually be carried out if the dispersion law is known in explicit form. In practice it is often just the dispersion law which is of basic interest. The calculation of field vectors according to formulae (33.1) – (33.4), which presents a major difficulty, can often be avoided. This fact represents the major advantage of using the Fourier integral method.

In conclusion we shall find the relation between the tensor ϵ_{ij} and the electrical conductivity tensor σ_{ij} .

Since the relation

$$\frac{\partial \mathbf{D}}{\partial t} = \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \mathbf{j}$$

is of universal character, then, substituting into it the expansions of the corresponding vectors in the Fourier integral and making use of the expression (33.13) and the definition of the electrical conductivity tensor we obtain

$$\epsilon_{ij} = \sigma_{ij} - i \frac{4\pi\sigma_{ij}}{\omega}. \quad (33.20)$$

This formula is a generalization of formula (31.12) which was obtained for the case where spatial dispersion is absent.

§34. The dispersion of light

We have discussed the effect of the scattering of electromagnetic waves by free and bound electrons in §36 of Part I. We shall now consider this effect from a somewhat different point of view. Namely, we shall calculate the

dielectric constant of a medium containing electrons which scatter the radiation.

As the simplest medium we shall consider a rarefied gas in which the polarization \mathbf{P} is equal to

$$\mathbf{P} = N\mathbf{d},$$

where N is the number of charges which produce the scattering, and \mathbf{d} is the dipole moment acquired by each of these. According to (36.2) of Part I, for \mathbf{d} we have

$$\mathbf{d} = \frac{e^2}{m} \frac{\mathbf{E}}{\omega_0^2 - \omega^2 + i\omega\gamma},$$

where ω_0 is the natural frequency of the oscillator.

The field \mathbf{E} in a rarefied gas is equal to the external field. Hence

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} = \mathbf{E} \left(1 + 4\pi N \frac{e^2}{m} \frac{1}{\omega_0^2 - \omega^2 + i\gamma\omega} \right), \quad (34.1)$$

and, consequently, making use of the definition (31.13), one can write for the dielectric constant

$$\epsilon = (n - i\kappa)^2 = 1 + \frac{4\pi e^2}{m} \frac{N}{\omega_0^2 - \omega^2 + i\gamma\omega}. \quad (34.2)$$

For a sufficiently rarefied gas the second term of (34.2), proportional to the number of electrons per unit volume, is small in absolute value, compared with unity, at all frequencies. Hence one can write approximately

$$n - i\kappa \approx 1 + \frac{2\pi e^2}{m} \frac{N}{\omega_0^2 - \omega^2 + i\gamma\omega}.$$

Separating the real and imaginary parts, we find

$$n = 1 + \frac{2\pi e^2 N}{m} \frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + \gamma^2 \omega^2}, \quad (34.3)$$

$$\kappa = \frac{2\pi e^2 N}{m} \frac{\gamma\omega}{(\omega_0^2 - \omega^2)^2 + \gamma^2 \omega^2}. \quad (34.4)$$

The formulae obtained above determine clearly measurable quantities — the refractive index and the absorption coefficient of the radiation. If we took a system of oscillators with natural frequencies $\omega_{01}, \omega_{02}, \dots, \omega_{0i}, \dots$, as a model of the gas, then in formulae (34.3) and (34.4) we would have to carry out the summation over different kinds of oscillators.

Let us discuss the dependence of n and κ on the frequency of the radiation in more detail. We assume, first of all, that ω differs sufficiently from the natural frequencies ω_{0i} . Then, neglecting $\gamma^2\omega^2$ in the denominator, we obtain

$$n^2 - \kappa^2 = 1 + \frac{4\pi e^2}{m} \sum_i \frac{N_i}{\omega_{0i}^2 - \omega^2}, \quad (34.5)$$

$$2n\kappa = \frac{4\pi e^2}{m} \sum_i \frac{N_i \omega \gamma}{(\omega_{0i}^2 - \omega^2)^2}. \quad (34.6)$$

We see that the absorption coefficient κ in the region $\omega \neq \omega_0$ is very small and that for n one can write

$$n \approx 1 + \frac{2\pi e^2}{m} \sum_i \frac{N_i}{\omega_{0i}^2 - \omega^2}. \quad (34.7)$$

Formula (34.7) expresses the dispersion law in a rarefied gas. The dependence of the refractive index on the frequency is shown in fig. IV.14. As

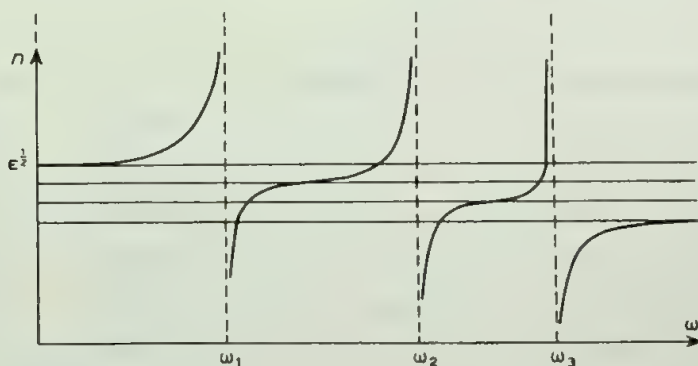


Fig. IV.14

ω approaches ω_0 , n increases sharply. For values of ω which are greater than ω_0 the index of refraction takes on a very small value (see below) and again increases with frequency. Each substance has its proper set of characteristic frequencies ω_{0i} . As an example we have in fig. IV.14 restricted ourselves to three frequencies.

We have already pointed out in Part I that real microscopic radiators, atoms and molecules, do not obey the laws of classical electrodynamics. Nevertheless, the harmonic oscillator, as a classical model of radiating systems, does yield a number of the basic properties of real emitters. This is seen particularly clearly in the example of dispersion. In quantum mechanics (see §108, Part V) it will be shown that the dispersion law is expressed by the formula

$$n = 1 + \frac{2\pi e^2 N}{m} \sum_r \frac{f_r}{\omega_r^2 - \omega^2}, \quad (34.8)$$

where the ω_r 's are the transition frequencies connected by the relation $\omega_r = (E_r - E_0)/\hbar$ between the r th and the lowest energy levels of the emitter, and the f_r 's are coefficients satisfying the condition $\sum f_r = 1$.

The resemblance between the accurate formula (34.8) and the formula based on the classical model (34.7) is obvious, although the quantities contained in the latter have a completely different meaning.

In the case of high frequencies satisfying the condition $\omega \gg \omega_r$, the quantum formula and the classical formula for the refractive index both assume the form

$$n = 1 - \frac{2\pi e^2}{m} \frac{N}{\omega^2}. \quad (34.9)$$

The refractive index turns out to be smaller than unity and to be independent of the properties of the substance, except for the number of electrons per unit volume. The meaning of this result is very simple: for the scattering of sufficiently hard radiation the binding of electrons in atoms ceases to play a significant part. Atomic electrons scatter the radiation just as free electrons. The refractive index n defined by formula (34.9) appears to be that of a medium which represents a gas of free electrons with density N . Formula (34.9) is applicable in the ultraviolet region to elements which are at the beginning of the periodic system and have a small binding energy for the electrons in the atoms, and it is applicable in the X-ray region to elements of the middle part and end of the periodic system.

A medium with $n < 1$ is optically less dense than vacuum. The phase velocity of electromagnetic waves in it is $v = c/n > c$ (see § 7 of Part II).

When sufficiently short electromagnetic waves are incident on the surface of a body, the phenomenon of total internal reflection (which will be described later, in § 35) can occur.

It is interesting to calculate the group velocity of electromagnetic waves in such a medium

$$v_g = \frac{d\omega}{dk} = \left(\frac{dk}{d\omega} \right)^{-1}.$$

We calculate $dk/d\omega$ making use of (34.9):

$$\frac{dk}{d\omega} = \frac{d}{d\omega} \frac{\omega}{v} = \frac{1}{c} \frac{d}{d\omega} (\omega n) = \frac{1}{c} \left(n + \omega \frac{dn}{d\omega} \right) = \frac{1}{c} \left(1 + \frac{2\pi e^2}{m} \frac{N}{\omega^2} \right).$$

Hence

$$v_g = \frac{c}{1 + \frac{2\pi e^2 N}{m\omega^2}} \approx c \left(1 - \frac{2\pi e^2 N}{m\omega^2} \right) = cn.$$

The group velocity, as is to be expected, turns out to be smaller than the velocity of light.

We now come back to the general case of dispersion and consider the frequency range $\omega \approx \omega_0$. In this case formula (34.5) is inadequate and one has to return to expressions (34.3) and (34.4). For $\omega \approx \omega_0$ one can write that

$$\omega_0^2 - \omega^2 \approx (\omega_0 - \omega) 2\omega_0.$$

We introduce a new variable

$$x = \frac{\omega - \omega_0}{\frac{1}{2}\gamma}.$$

Then

$$n = 1 - \frac{2\pi e^2 N}{m\gamma\omega_0} \frac{x}{1 + x^2},$$

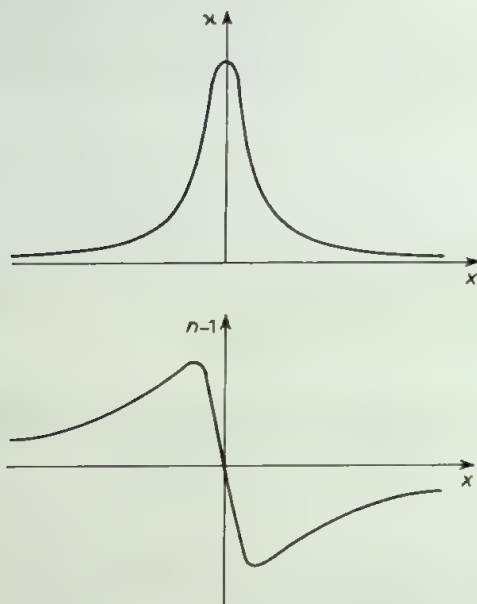


Fig. IV.15

$$\kappa = \frac{2\pi e^2 N}{m\gamma\omega_0} \frac{1}{1+x^2}.$$

The curves $\kappa(x)$ and $n(x) - 1$ are shown in fig. IV.15. We see that maximum absorption occurs for $x = 0$ (i.e. $\omega = \omega_0$). This absorption or, more precisely, attenuation, is associated with a large value of the cross-section for scattering (see §36 of Part I). It decreases sharply with increasing $|x|$, reaching half of its maximum value for $x = \pm 1$. The refractive index in the region $-1 < x < 1$ displays a dependence on x which is completely different from that for $|x| > 1$ (shown in fig. IV.15). For $|x| > 1$, n increases with increasing x , while for $x = \mp 1$ it reaches respectively its maximum and minimum values. In the interval $-1 < x < 1$, $n(x)$ is not an increasing but a decreasing function of x , reaching the value $n = 1$ at $x = 0$. This region, in which the index of refraction decreases with increasing frequency, is called the region of anomalous dispersion. In the region of anomalous dispersion a rarefied gas possesses its highest absorption (minimum transparency).

We shall come back to the problems of dispersion in quantum mechanics where it will be shown that besides scattering without a change in frequency scattering with a change in frequency (so-called Raman scattering) is possible.

§35. Geometrical optics

In the above we have found the law of the propagation of plane electromagnetic waves in homogeneous transparent media. A characteristic of a plane wave is the fact that its equiphase surface is infinitely large and plane.

Plane electromagnetic waves are particularly simple because their amplitude and wave vector remain constant in space and time. It turns out that, under certain conditions and with a certain degree of accuracy, analogous properties are possessed by arbitrary electromagnetic waves. As will be shown below, it is necessary for this that the curvature of the wave surface is sufficiently small over spatial regions which are large compared with the wavelength.

The approximate replacement of the wave surface by a plane is of very great practical importance, since in the case of an arbitrary form of the surface the laws of propagation of the waves turn out to be very complex.

From the aforesaid it is clear that such a substitution is allowed, in any case, if the wavelength is very small. Therefore the limiting case of electromagnetic waves whose wavelength $\lambda \rightarrow 0$ deserves particular consideration.

In practice the region of visible light already corresponds to this case. Indeed, the wavelength of visible light is about 5×10^{-5} cm. It is always very small in comparison with the size of macroscopic bodies and optical devices. For this reason the study of the laws of propagation of electromagnetic waves in the limiting case $\lambda \rightarrow 0$ is called geometrical optics.

We shall below, for the same reason, sometimes call an electromagnetic wave a light wave, and we shall speak of the propagation of light. This should not lead to misunderstanding.

In a transparent medium each of the components of the field vectors \mathbf{E} and \mathbf{H} satisfies the wave equation

$$\nabla^2 F - \frac{\epsilon}{c^2} \frac{\partial^2 F}{\partial t^2} = 0.$$

Here F denotes an arbitrary component of the field.

In a bounded medium, in addition to the wave equation the components of the field vectors must also satisfy the boundary conditions. Therefore the

components of the vectors \mathbf{E} and \mathbf{H} are connected with each other by certain relations and are not quite independent. However, at a sufficiently large distance from the interface of media this fact can be disregarded and all the components of \mathbf{E} and \mathbf{H} can be considered as independent. Then the wave field can be characterized by the scalar quantity F .

Considering a monochromatic wave, we can write

$$F = f(x, y, z) e^{-i\omega t}.$$

The function f obviously satisfies the equation

$$\nabla^2 f + \frac{\epsilon\omega^2}{c^2} f = 0.$$

Introducing the wave number $k = \omega/c$ and the refraction index $n = \epsilon^{1/2}$, one can rewrite the above equation in the form

$$\nabla^2 f + n^2 k^2 f = 0. \quad (35.1)$$

To a plane wave there corresponds the solution of eq. (35.1)

$$f = a e^{in(\mathbf{k} \cdot \mathbf{r})} a e^{ik(n\mathbf{k}_0 \cdot \mathbf{r})} = a e^{ikl}, \quad (35.2)$$

where a is a constant amplitude, and $l = n(\mathbf{k}_0 \cdot \mathbf{r})$ is a quantity which is usually called the optical path length and is a linear function of coordinates. It differs from the geometrical path by the factor n .

We have to find the solutions of eq. (35.1) in a general form, without confining ourselves to a plane wave. However, we shall consider the limiting case $\lambda \rightarrow 0$ or $k \rightarrow \infty$. We shall try to write the solution of eq. (35.1) in the limiting case of very large k as

$$f = a e^{i\psi(x, y, z)}. \quad (35.3)$$

Here the quantity ψ , called the eikonal, has a form which approaches as closely as possible that of the phase of the plane wave

$$\psi = k\mathcal{L}(x, y, z),$$

where $\mathcal{L}(x, y, z)$ is a function of the coordinates which is sufficiently close to a linear function. Thus, we shall try to find the solution of (35.1) in the form

$$f = a e^{ik\mathcal{L}(x,y,z)} . \quad (35.4)$$

We calculate the derivatives of the function f :

$$\nabla f = a \nabla e^{ik\mathcal{L}} = ika e^{ik\mathcal{L}} \nabla \mathcal{L}$$

$$\nabla^2 f = \nabla \cdot \nabla a e^{ik\mathcal{L}} = ika e^{ik\mathcal{L}} \nabla^2 \mathcal{L} - k^2 a e^{ik\mathcal{L}} (\nabla \mathcal{L})^2 .$$

Hence eq. (35.1) assumes the form

$$ika e^{ik\mathcal{L}} \nabla^2 \mathcal{L} - k^2 a e^{ik\mathcal{L}} (\nabla \mathcal{L})^2 + n^2 k^2 a e^{ik\mathcal{L}} = 0 ,$$

or

$$(\nabla \mathcal{L})^2 - (i/k) \nabla^2 \mathcal{L} = n^2 . \quad (35.5)$$

The quantity \mathcal{L} does not depend on k . Therefore, if for $k \rightarrow \infty$ the inequality

$$|\nabla \mathcal{L}|^2 \gg k^{-1} \nabla^2 \mathcal{L} , \quad (35.6)$$

is fulfilled, then one can pass to the limit $k \rightarrow \infty$ in eq. (35.5) and drop the second term. Then eq. (35.5) assumes the form

$$|\nabla \mathcal{L}|^2 = n^2 ,$$

or

$$(\nabla \psi)^2 = n^2 k^2 . \quad (35.7)$$

In the coordinate representation this last equation has the form

$$\left(\frac{\partial \psi}{\partial x} \right)^2 + \left(\frac{\partial \psi}{\partial y} \right)^2 + \left(\frac{\partial \psi}{\partial z} \right)^2 = n^2 k^2 . \quad (35.7')$$

Eq. (35.7) is called the eikonal equation. In deriving it we have not made any assumptions about the values of the refractive index n , which can be an arbitrary function of the coordinates.

Knowing the solutions of eq. (35.7), we find an approximate solution of the wave equation (35.1) with the assumption that k is a sufficiently large number. In the limit $k \rightarrow \infty$ this solution is the same as a plane wave propagat-

ing in the direction of n . The condition of applicability of eq. (35.7) is (35.6). Later we shall discuss the question of the physical conditions to which this inequality corresponds and when it may not be satisfied.

If ψ is a solution of the eikonal equation, then

$$\psi = k\mathcal{L}(x, y, z) = \text{const}$$

is the equation of an equiphase surface. The propagation of electromagnetic waves takes place in the direction of the normal to the constant-eikonal surface, i.e. in the direction of the vector $\nabla\mathcal{L}$. This direction of propagation of waves is called the light ray direction.

Over a small region of space the eikonal ψ can be expanded in a series, writing

$$\psi \approx \frac{\partial\psi}{\partial\mathbf{r}} \cdot \mathbf{r} = k \frac{\partial\mathcal{L}}{\partial\mathbf{r}} \cdot \mathbf{r} = k(\nabla\mathcal{L}) \cdot \mathbf{r}. \quad (35.8)$$

For f , according to (35.3), we have

$$f = a e^{ik(\nabla\mathcal{L}) \cdot \mathbf{r}} = a e^{ikl}, \quad (35.9)$$

The optical path l turns out to be equal to $(\nabla\mathcal{L}) \cdot \mathbf{r}$.

Comparing (35.9) with (35.2), we see that, within the framework of the applicability of geometrical optics, over a small region of space each wave can be considered as a plane wave with its wave vector directed along the normal to the constant-eikonal surface. In particular, for a homogeneous medium with a constant refractive index, it follows from (35.7) that

$$\nabla\psi = \text{const}.$$

This means that adjacent equi-eikonal surfaces are at equal distances from each other. The normals to these surfaces are straight lines. This statement has a simple meaning: in a homogeneous medium light rays are propagated rectilinearly.

We shall not dwell on the application of geometrical optics to optical and electro-optical systems*. We shall discuss only the question of the meaning

* See, for example, L.D.Landau and E.M.Lifshitz, *The classical theory of fields* (Pergamon Press, London, 1962) §55–57; M.Born and E.Wolf, *Principles of optics* (Pergamon Press, London, 1959, 1965).

of the inequality (35.6) which serves as the condition of applicability of the approximation of geometrical optics.

As is known from geometry, the quantity $\nabla^2\psi$ determines the mean radius of curvature of a slightly curved surface $\psi = \text{const}$. Thus, the inequality (35.6) has a simple geometrical meaning: it means that the mean radius of curvature of the constant-eikonal surface must be large in comparison with the wavelength λ . In other words, the light perturbation must vary smoothly from point to point. The conditions of applicability of the approximation of geometrical optics are undoubtedly violated in the immediate vicinity of the emitter. Here the variation of the wave field from point to point is very rapid. The quantity $\nabla^2\psi$ can be of the same order of magnitude as the wavelength.

Consider a second important case where it turns out that it is impossible to make use of the concepts of geometrical optics.

Let a completely non-transparent body, or screen, be placed in the path of light rays. In the approximation of geometrical optics the light rays do not penetrate the region behind the screen (fig. IV.16). Behind the opaque screen there is a region of shadow.

Let us consider the boundary separating the illuminated region and the shadow (straight lines AB and CD in fig. IV.16). Here a discontinuous change in the amplitude from the value a to zero occurs. Hence the derivatives of f with respect to the coordinates become infinite at the boundary between light and shadow. Here the conditions for applicability of geometrical optics are again violated. In the next paragraph we shall discuss physical phenomena which arise in this case.

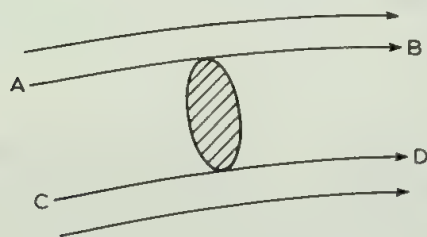


Fig. IV.16

§36. Diffraction

As we have just shown, at the boundary separating the illuminated and the shaded regions the approximation of geometrical optics and the concept of rectilinear light rays becomes inadmissible.

Naturally, there cannot exist a sharp boundary between these regions. Near the edge of the non-transparent screen the propagation of light cannot in reality be described by a simplified wave equation. The solution of the correct equation, taking into account the boundary conditions at the surface of the screen, leads to an intensity distribution corresponding to a gradual transition from the region of total illumination to that of shadow. At the boundary between the illuminated and shaded regions departure from the rectilinear propagation of light occurs, as if the rays were bent. This is called diffraction.

The solution of the boundary value problem of diffraction turns out to be very complicated. Therefore special methods for a simplified calculation of diffraction phenomena have been developed. We shall restrict ourselves to the discussion of the most important and at the same time simplest case.

Let a plane parallel beam of light be incident on a screen having an opening of arbitrary shape. The light beam comes from a source of intensity I_0 which is distant from the screen. At the edge of the screen diffraction of the light occurs. We are interested in the light intensity distribution at a large distance from the opening (large compared with the size of the opening). Diffraction in the case where both the light source and the point of observation are located at infinitely large distances from the aperture which produces the diffraction, or from the screen, is called Fraunhofer diffraction.

For Fraunhofer diffraction the waves which fall on the screen and those which pass through the opening at a large distance from its edge can be assumed to be plane waves. We shall call the wave incident on the screen the primary wave. This wave, falling on the opaque screen, is completely absorbed. However, part of it passes through the opening in the screen. The opening in the screen becomes an emitter of electromagnetic waves into the space behind the screen.

Each point of the area of the opening is a source of secondary spherical waves which reach the point of observation with a certain retardation.

If the value of any component of the field at a point \mathbf{r}_0 located in the area of the opening is $f_0 = \text{const } e^{-i\omega t}$, then the perturbation coming from this point to the point of observation N will be of the order of magnitude of $R^{-1} f_0(\tau) = \text{const } e^{-i\omega\tau}$, where $\tau = t - R/c$ is the delay time (see §23 of Part I).

The total field at the point of observation is obtained on the basis of the principle of superposition by summing over all the emitters, i.e. over all points of the opening in the screen.

Thus, at the point of observation N we have

$$\begin{aligned} f_N &= \text{const} \cdot \int \frac{e^{-i\omega(t-R/c)}}{R} dS = \text{const} \cdot e^{-i\omega t} \int \frac{e^{i\omega R/c}}{R} dS = \\ &= \text{const} \cdot e^{-i\omega t} \int \frac{e^{ikR}}{R} dS. \end{aligned} \quad (36.1)$$

Here f denotes any component of the electric or magnetic field. In the same approximation as in the preceding section we ignore polarization phenomena and assume all components of the field to be independent.

Formula (36.1) is of a general character. In calculating the integral (36.1) we shall bear in mind that the distance to the point of observation is large in comparison with the size of the opening (Fraunhofer diffraction!).

We introduce an origin, locating it at a point of the opening (fig. IV.17). Then one can, as usual (see, for example, §15 or §26 of Part I), write

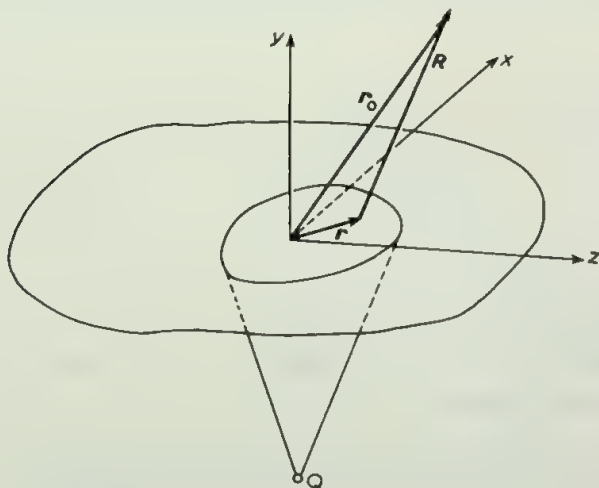


Fig. IV.17

$$R = |\mathbf{r}_0 - \mathbf{r}| \approx r_0 - \mathbf{n}_0 \cdot \mathbf{r},$$

where $\mathbf{n}_0 = \mathbf{r}_0/r_0$ is the unit vector in the direction of \mathbf{r}_0 .

The quantity $|\mathbf{r}|$, which varies within the limits of the opening, is small in comparison with $|\mathbf{r}_0|$, but may be large compared with the wavelength. Therefore the term $\mathbf{n}_0 \cdot \mathbf{r}$ can be dropped in the denominator of formula (36.1) but must be retained in the exponent. Hence we have

$$f \approx \frac{\text{const}}{r_0} \exp[-i\omega t] \exp[ikr_0] \int \exp[-ik\mathbf{n}_0 \cdot \mathbf{r}] dS = A \int \exp[-i\mathbf{q} \cdot \mathbf{r}] dS, \quad (36.2)$$

where $\mathbf{q} = k\mathbf{n}_0$, and A denotes the whole set of factors in front of the integral. The value of this constant will be expressed in terms of the intensity of the incident wave.

Choosing the plane of the opening to be the plane $z = 0$, one can write the quantity $\mathbf{q} \cdot \mathbf{r}$ in the form

$$\mathbf{q} \cdot \mathbf{r} = q_1 x + q_2 y = k \frac{x_0}{r_0} x + k \frac{y_0}{r_0} y = kx \cos \alpha + ky \cos \beta. \quad (36.3)$$

Here $\cos \alpha$ and $\cos \beta$ are the direction cosines of the vector \mathbf{n}_0 .

If the wave vector of the incident wave is normal to the plane of the opening and the angle of deflection is small, then it is easily seen that

$$\cos \alpha \approx \sin \theta_1 \approx \theta_1,$$

$$\cos \beta \approx \sin \theta_2 \approx \theta_2,$$

where θ_1 and θ_2 are the angles of diffraction in the xz -plane and yz -plane. In this case

$$q_1 \approx k\theta_1, \quad q_2 \approx k\theta_2. \quad (36.4)$$

The integration in (36.2) is carried out over the area of the opening. If one introduces the function

$$\xi(x, y) = \begin{cases} 1 & \text{for } (x, y) \text{ within the limits of the opening} \\ 0 & \text{for } (x, y) \text{ outside the limits of the opening,} \end{cases}$$

then one can write (36.2) in the form

$$f = A \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \xi(x, y) \exp [-i(q_1 x + q_2 y)] dx dy \quad (36.5)$$

where the integration is carried out over the entire area of the screen (including the opening).

In order to determine the constant A we note that (36.5) represents the Fourier integral of the function $\xi(x, y)$. Hence

$$\xi(x, y) = \frac{1}{(2\pi)^2 A} \int \exp [i(q_1 x + q_2 y)] f(q_1, q_2) dq_1 dq_2 .$$

From formula (II.9) it follows that

$$\int |\xi(x, y)|^2 dx dy = \frac{1}{(2\pi)^2 A^2} \int |f(q_1, q_2)|^2 dq_1 dq_2 .$$

By the definition of ξ the integral on the left is equal to unity multiplied by the area S of the opening. The integral $\int |f(q_1, q_2)|^2 dq_1 dq_2$ by definition represents the total intensity behind the screen. The latter is equal, obviously, to the total intensity of light passing through the opening in the screen, i.e. to

$$\int |f|^2 dq_1 dq_2 = I_0 S ,$$

where I_0 is the intensity of the incident light. Hence we find the value of A :

$$A = I_0^{1/2} / 2\pi \quad (36.6)$$

and, finally, we have

$$f(q_1, q_2) = (I_0^{1/2} / 2\pi) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp [-i(q_1 x + q_2 y)] \xi(x, y) dx dy . \quad (36.7)$$

Formula (36.7) gives the general solution of the problem posed; the distribution of the wave field behind a screen having an aperture at a large distance from it.

The intensity distribution is given by the quantity

$$dI = |f(q_1, q_2)|^2 dq_1 dq_2. \quad (36.8)$$

From the meaning of the quantities q_1 and q_2 it is clear that $|f|^2 dq_1 dq_2$ represents the intensity of the diffracted light for given values of the diffraction angles θ_1 and θ_2 :

$$dI = |f(q_1, q_2)|^2 k^2 d\Omega. \quad (36.8')$$

Up to now we have considered diffraction from a small opening in a screen. We now assume that light falls on a small screen whose size and form are the same as those of the opening in the screen previously considered. Such a screen is called a complementary screen with respect to the opening. The diffraction at the edge of the complementary screen is easily found from the following considerations. We match the complementary screen with the opening. Then behind the new screen, which is continuous, there will be no field of light. This latter fact can be expressed as follows: as a result of the superposition of the fields diffracted from the opening and from the complementary screen the total field is reduced to zero. Hence it follows that the field behind the complementary screen can be connected with the field behind the opening by the relation

$$f_{\text{compl}} = -f,$$

or

$$|f_{\text{compl}}|^2 dq_1 dq_2 = |f|^2 dq_1 dq_2. \quad (36.9)$$

The opening and its complementary screen give identical intensity distributions of the diffracted light. This proposition is called Babinet's principle.

We shall apply the general expression for the intensity distribution to three cases which are of great practical importance.

The first of these is the diffraction at an infinitely long slit in an opaque screen. Let the width of the slit be a . We locate the origin in the middle of the slit, and direct the y -axis parallel to its edge. The direction of the incident light coincides with the direction of the z -axis. It is obvious that the diffraction comes only from the two edges of the slit, i.e. in the integral (36.7) one has to integrate only with respect to the coordinate x . For an infinitely long slit, no diffraction of rays falling on the screen occurs in the direction of y . Hence we shall refer all quantities to unit length of the slit. In the one-dimensional case a calculation analogous to that which has been carried

out before gives the value $A = (I_0/2\pi)^{\frac{1}{2}}$. Formula (36.7) in the one-dimensional case has the form

$$f = (I_0/2\pi)^{\frac{1}{2}} \int_{-\frac{1}{2}a}^{\frac{1}{2}a} e^{-iqx} dx = (2I_0/\pi)^{\frac{1}{2}} \frac{\sin \frac{1}{2}aq}{q} = (2I_0/\pi)^{\frac{1}{2}} \frac{\sin \frac{1}{2}ka\theta}{k\theta}. \quad (36.10)$$

The intensity of light diffracted at an angle between θ and $\theta + d\theta$ is equal to

$$dI = I_0 \frac{2 \sin^2 \frac{1}{2}ka\theta}{\pi(k\theta)^2} dq = \frac{I_0 ka^2}{2\pi} \frac{\sin^2 \frac{1}{2}ka\theta}{(\frac{1}{2}ka\theta)^2} d\theta. \quad (36.11)$$

The quantity $I_0 a$ represents the total intensity of light incident on unit length of the slit.

As a result of diffraction, a system of light and dark bands arises parallel to the slit.

The intensity distribution as a function of $x = \frac{1}{2}ka\theta$ is shown in fig. IV.18. In the direction $\theta = 0$ there is a principal maximum of the intensity. At points

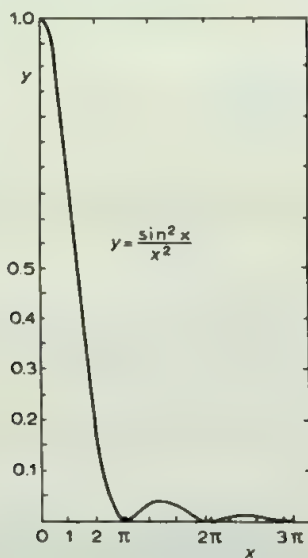


Fig. IV.18

$\frac{1}{2}ka\theta = \frac{1}{2}(n+1)\pi$ ($n=1,2,\dots$) the intensity has a number of secondary maxima. In these maxima the intensity is considerably lower than in the principal maximum, and decreases with increasing order of the maximum (the number n).

One can in an analogous way find the diffraction from a circular opening of radius a in the screen. We locate the origin at the centre of the circle. Then

$$\begin{aligned} f &= (I_0^{\frac{1}{2}}/2\pi) \int_0^a \int_0^{2\pi} e^{-i\mathbf{q}\cdot\mathbf{r}} r \, dr d\varphi = \\ &= (I_0^{\frac{1}{2}}/2\pi) \int_0^a \int_0^{2\pi} e^{-iqr \cos \varphi} r \, dr d\varphi = (I_0^{\frac{1}{2}}/2\pi) \int_0^a r \, dr \int_0^{2\pi} e^{-iqr \cos \varphi} d\varphi. \end{aligned}$$

From the theory of Bessel functions it is known that

$$\int_0^{2\pi} e^{-iqr \cos \varphi} d\varphi = 2\pi J_0(qr),$$

where J_0 is the Bessel function of zero order *. Hence

$$f = I_0^{\frac{1}{2}} \int_0^a J_0(qr) r \, dr.$$

In the theory of Bessel functions it is proved that

$$\int_0^a J_0(qr) r \, dr = \frac{a}{q} J_1(qa).$$

Hence, finally,

$$f = I_0^{\frac{1}{2}} \frac{a}{q} J_1(aq). \quad (36.12)$$

* This formula is most simply obtained by expanding the exponent in a series and integrating term by term. See V.I. Smirnov, *Course of higher mathematics*, Vol. III, Part 2 (Pergamon Press, Oxford, 1964).

The intensity of radiation diffracted into a solid angle $d\Omega$ is equal to

$$dI = |f|^2 k^2 d\Omega = I_0 a^2 \frac{J_1^2(ka\theta)}{\theta^2} d\Omega = (I_0 a^2 \pi) \frac{J_1^2(ka\theta)}{\pi\theta^2} d\Omega, \quad (36.13)$$

where $I_0 \pi a^2$ is the total intensity of radiation incident on the opening.

As a result of diffraction at a circular opening, a system of concentric circles of illumination arises.

The angular distribution of the diffracted radiation given by the function $[J_1^2(ka\theta)]/\pi\theta^2$ is very similar to the distribution shown in fig. IV.18. The height of the principal maximum is relatively greater, but the spacings between secondary maxima are close to those which are shown in fig. IV.18.

It should be recalled that identical formulae give the intensity distribution for diffraction from the complementary screens. Naturally there correspond to the intensity maxima for the diffraction from the aperture the maxima for the diffraction from the complementary opaque screen.

In conclusion we shall dwell on the diffraction which occurs when light is passing through a set of N infinitely long thin slits which are located at equal distances from each other (a diffraction grating).

Let the distance between the centres of neighbouring slits be d , and let the grating be oriented along the x -axis. The coordinates of the centres of the slits are $x_n = nd$, where the n are integers $n = 0, 1, 2, \dots, N$. Then (36.7) gives

$$f = (I_0/2\pi)^{\frac{1}{2}} \sum_{n=0}^N e^{-iqx_n} \int_{-\frac{1}{2}a}^{\frac{1}{2}a} e^{-iqx} dx = f_1 \sum_{n=0}^N e^{-iqnd} = f_1 \frac{1 - e^{-iNdq}}{1 - e^{-idq}},$$

where f_1 is defined by formula (36.10).

The intensity of radiation diffracted from the system of slits is given by the formula

$$dI = |f|^2 dq \sim I_0 \left(\frac{\sin \frac{1}{2} Nqd}{\sin \frac{1}{2} qd} \right)^2 \left(\frac{\sin \frac{1}{2} qa}{\frac{1}{2} qa} \right)^2 dq. \quad (36.14)$$

The interference pattern which arises corresponds to the superposition of the diffraction from different slits, but not to a simple sum of intensities.

An important part is played in modern physics by the phenomenon of diffraction of X-rays (as well as electrons, neutrons etc.) by crystals (see Part V).

We shall discuss below the simplest theory of X-ray diffraction by crystals.

Let a crystal lattice be formed by a set of atoms of the same kind (i.e. atoms which are identical in their properties). We assume that absorption in the lattice is absent.

We further assume that an X-ray beam propagating in the direction \mathbf{n}_1 falls on the lattice. The coordinate of the point of the crystal at which the m th atom of the lattice is located can be written in the form

$$\mathbf{r}_m = l\mathbf{a} + n\mathbf{b} + p\mathbf{c}, \quad (36.15)$$

where l, n, p are integers (positive and negative), and the vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are the unit vectors of the lattice (for more detail see §109 of Part V; we restrict ourselves here to the case of the so-called primitive lattice).

A wave with amplitude

$$f = f_0 \exp \left[-i \left(\omega t - \frac{\omega \mathbf{r}_m \cdot \mathbf{n}_1}{c} \right) \right],$$

where f is an arbitrary component of the field vectors and f_0 is a constant, arrives at the point \mathbf{r}_m . The atom which is located at the point \mathbf{r}_m scatters the wave, and a wave with an amplitude

$$f_N = \frac{f_0}{R_m} \exp \left[-i \left(\omega t - \frac{\omega \mathbf{r}_m \cdot \mathbf{n}_1}{c} - \frac{\omega R_m}{c} \right) \right]$$

arrives at the point of observation N which is located at a distance R_m from the m th atom.

Assuming that

$$R_m \approx r_0 - \mathbf{n} \cdot \mathbf{r}_m,$$

where r_0 is the distance from the point N to the origin, we find for the amplitude of the wave scattered by the m th atom

$$f_N = f_0 r_0^{-1} \exp [-i\omega t] \exp [ikr_0] \exp [-ik\{\mathbf{r}_m \cdot (\mathbf{n} - \mathbf{n}_1)\}].$$

If one disregards the secondary action of the wave scattered by the m th atom on other atoms of the lattice, then one can write the total intensity at the point of observation in the form

$$I \sim I_0 \left| \sum \exp [-ik\{\mathbf{r}_m \cdot (\mathbf{n} - \mathbf{n}_1)\}] \right|^2, \quad (36.16)$$

where the summation is carried out over all the atoms of the crystal (the numbers l, n, p).

Taking into account the expression (36.15), one can write

$$\begin{aligned} k[\mathbf{r}_m \cdot (\mathbf{n} - \mathbf{n}_1)] &= kl[\mathbf{a} \cdot (\mathbf{n} - \mathbf{n}_1)] + kn[\mathbf{b} \cdot (\mathbf{n} - \mathbf{n}_1)] + kp[\mathbf{c} \cdot (\mathbf{n} - \mathbf{n}_1)] = \\ &= l[\mathbf{a} \cdot (\mathbf{k} - \mathbf{k}_1)] + n[\mathbf{b} \cdot (\mathbf{k} - \mathbf{k}_1)] + p[\mathbf{c} \cdot (\mathbf{k} - \mathbf{k}_1)]. \end{aligned}$$

Then each of the sums can be calculated directly. For example,

$$\sum_{l=0}^{N_1} \exp [-il\{\mathbf{a} \cdot (\mathbf{k} - \mathbf{k}_1)\}] = \frac{1 - \exp [-iN_1\{\mathbf{a} \cdot (\mathbf{k} - \mathbf{k}_1)\}]}{1 - \exp [-i\{\mathbf{a} \cdot (\mathbf{k} - \mathbf{k}_1)\}]} ,$$

where N_1 is the number of lattice points along the first rib of the crystal. Hence it follows immediately that

$$I \sim I_0 \frac{\sin^2 \frac{1}{2} N_1 [\mathbf{a} \cdot (\mathbf{k} - \mathbf{k}_1)]}{\sin^2 \frac{1}{2} [\mathbf{a} \cdot (\mathbf{k} - \mathbf{k}_1)]} \frac{\sin^2 \frac{1}{2} N_2 [\mathbf{b} \cdot (\mathbf{k} - \mathbf{k}_1)]}{\sin^2 \frac{1}{2} [\mathbf{b} \cdot (\mathbf{k} - \mathbf{k}_1)]} \frac{\sin^2 \frac{1}{2} N_3 [\mathbf{c} \cdot (\mathbf{k} - \mathbf{k}_1)]}{\sin^2 \frac{1}{2} [\mathbf{c} \cdot (\mathbf{k} - \mathbf{k}_1)]} . \quad (36.17)$$

Formula (36.7) is called the Laue formula. It is a generalization of the formula of the diffraction grating to the case of three dimensions with unequal spacings.

The principal maxima lie in directions determined by a set of equations which are called Bragg's conditions:

$$\begin{aligned} \mathbf{a} \cdot (\mathbf{n} - \mathbf{n}_1) &= 2\pi h_1/k = h_1 \lambda, \\ \mathbf{b} \cdot (\mathbf{n} - \mathbf{n}_1) &= 2\pi h_2/k = h_2 \lambda, \\ \mathbf{c} \cdot (\mathbf{n} - \mathbf{n}_1) &= 2\pi h_3/k = h_3 \lambda, \end{aligned} \quad (36.18)$$

where h_1, h_2, h_3 are integers.

Each of these expressions is the condition for the reduction to zero of the corresponding factor in the denominator of formula (36.17).

It should be emphasized that for given wavelength λ and direction of the

incident beam n_1 the set of three equations (36.18) allows the determination of two independent components of the unit vector \mathbf{n} . The system of eqs. (36.18) has a solution only for a definite wavelength. The reflection of X-rays from the crystal is said to have a selective character. The difference between one-dimensional and two-dimensional diffraction gratings and crystals lies in this fact.

In practice, when the wavelength which can be selectively reflected from the lattice is unknown, then the crystal is either irradiated by a continuous X-ray spectrum or the orientation of the diffraction crystal (the direction of n_1) is changed. In both cases the crystal itself "selects" conditions (the wavelength or orientation) under which the selective reflection is possible, i.e. under which the equations (36.18) are simultaneously fulfilled.

§37. The reflection and refraction of electromagnetic waves at the boundary between media

Up to now we have studied the propagation of electromagnetic waves in a homogeneous medium. We shall now consider the electromagnetic wave incident on the boundary between two dielectric media which differ from each other in the values of their dielectric constants. For simplicity of calculation we shall assume that the magnetic permeabilities of the two media are equal to unity.

We assume the boundary to be plane. It is clear that near the boundary the physical properties of the media may differ from those within their volumes, so that a thin transition layer exists at the boundary. However, if the thickness of the latter is small compared with the wavelength, its effect can be neglected and the boundary can be considered as a mathematical interface having no thickness.

We assume that a monochromatic plane wave is incident on this surface, which we choose as the plane $z = 0$. For simplicity we assume that the vector \mathbf{k} characterizing the direction of propagation of the incident wave lies in the xz -plane. We shall call the medium in which the incident wave is propagating the first medium.

We write the vectors of the electric and magnetic fields of the incident wave in the first medium in the form

$$\mathbf{E}^{\text{inc}} = A \exp [i(\omega t - \mathbf{k} \cdot \mathbf{r})] = A \exp [i(\omega t - kx \cos \alpha - kz \cos \gamma)] ,$$

$$\mathbf{H}^{\text{inc}} = C \exp [i(\omega t - \mathbf{k} \cdot \mathbf{r})] = C \exp [i(\omega t - kx \cos \alpha - kz \cos \gamma)] ,$$

where $\cos \alpha$ and $\cos \gamma$ are the direction cosines of the wave vector. Since the wave is lying in the xz -plane, $\cos \beta = 0$. The wave number k is connected with the frequency by the relation

$$k = \omega \epsilon_1^{\frac{1}{2}} / c .$$

At the interface the vectors \mathbf{E}^{inc} and \mathbf{H}^{inc} must satisfy the boundary conditions (5.2) and (5.4). The quantities \mathbf{E} and \mathbf{H} are not independent but are perpendicular to each other, and their absolute values are related by the expression (31.31). Hence two independent boundary conditions can be satisfied simultaneously only by assuming that the incident wave is partially transmitted into the second medium and partially reflected from the interface. We shall call the wave transmitted into the second medium the refracted wave, because its direction of propagation does not, as a rule, coincide with the direction of propagation of the incident wave. Let

$$\mathbf{E}^{\text{refl}} = \mathbf{A}^{\text{refl}} \exp [i(\omega^{\text{refl}} t - \mathbf{k}^{\text{refl}} \cdot \mathbf{r})] ,$$

$$\mathbf{E}^{\text{refr}} = \mathbf{A}^{\text{refr}} \exp [i(\omega^{\text{refr}} t - \mathbf{k}^{\text{refr}} \cdot \mathbf{r})] .$$

denote the electric vectors of the reflected and refracted waves. The total field in the first medium is characterized by the vector

$$\mathbf{E}_1 = \mathbf{E} + \mathbf{E}^{\text{refl}} = \mathbf{A} \exp [i(\omega t - \mathbf{k} \cdot \mathbf{r})] + \mathbf{A}^{\text{refl}} \exp [i(\omega^{\text{refl}} t - \mathbf{k}^{\text{refl}} \cdot \mathbf{r})] .$$

The boundary condition (5.2) at the interface of two media gives

$$\begin{aligned} \mathbf{A}_{\text{tg}}^{\text{inc}} \exp [i(\omega t - kx \cos \alpha)] + \mathbf{A}_{\text{tg}}^{\text{refl}} \exp [i(\omega^{\text{refl}} t - k^{\text{refl}} x \cos \alpha^{\text{refl}} - k^{\text{refl}} y \cos \beta^{\text{refl}})] \\ = \mathbf{A}_{\text{tg}}^{\text{refr}} \exp [i(\omega^{\text{refr}} t - k^{\text{refr}} x \cos \alpha^{\text{refr}} - k^{\text{refr}} y \cos \beta^{\text{refr}})] . \end{aligned}$$

This equation must hold for arbitrary values of time t and coordinates x and y . This is possible only if

$$\omega = \omega^{\text{refl}} = \omega^{\text{refr}} , \quad (37.1)$$

$$k \cos \alpha = k^{\text{refl}} \cos \alpha^{\text{refl}} = k^{\text{refr}} \cos \alpha^{\text{refr}} ,$$

$$\cos \beta^{\text{refl}} = \cos \beta^{\text{refr}} = 0 .$$

The first condition means that the reflection and refraction occur without any change in the frequency. The third equality shows that the reflected and refracted waves lie in the same plane as the incident wave. Finally, taking into account that $k = \omega/v = \epsilon^{1/2} \omega/c$, one can rewrite the second relation in the form

$$(\epsilon^{1/2} \omega/c) \cos \alpha = (\epsilon^{1/2} \omega/c) \cos \alpha^{\text{refl}},$$

$$(\epsilon^{1/2} \omega/c) \cos \alpha = (\epsilon^{1/2} \omega/c) \cos \alpha^{\text{refr}},$$

whence we find that

$$\alpha = \alpha^{\text{refl}}, \quad (37.2)$$

i.e. that the angle of incidence is equal to the angle of reflection, and

$$\frac{\cos \alpha}{\cos \alpha^{\text{refr}}} = \left(\frac{\epsilon_2}{\epsilon_1} \right)^{1/2}. \quad (37.3)$$

Instead of the direction cosines $\cos \alpha$ and $\cos \alpha^{\text{refr}}$ one usually introduces $\sin \theta$ and $\sin \theta^{\text{refr}}$, where θ and θ^{refr} are the angles formed by the incident and refracted waves with the normal to the plane $z = 0$, and which are called the angle of incidence (θ) and the angle of refraction (θ^{refr}). For the latter we find

$$\frac{\sin \theta}{\sin \theta^{\text{refr}}} = n_{12} = \left(\frac{\epsilon_2}{\epsilon_1} \right)^{1/2}. \quad (37.4)$$

Here n_{12} denotes the refractive index of the boundary of the second medium with respect to the first medium.

We arrive at the laws of reflection and refraction of light. However, the value of the refractive index turns out to be related to the dielectric constants of the media. In particular, if the first medium is vacuum so that $\epsilon_1 = 1$, then

$$n = \epsilon_0^{1/2}$$

is called the refractive index of the given medium.

The expression for the refractive index justifies the terminology which we have introduced in §33.

In considering the phenomena of reflection and refraction of electromagnetic waves, which had been extensively investigated for light even before the

establishment of its electromagnetic nature, the optical terminology is adopted. A medium with a higher refractive index is said to be optically more dense than a medium with a lower refractive index.

Having established the directions of propagation of the reflected and refracted waves, we go on to the calculation of their amplitudes. Taking into account eqs. (37.1) we obtain, using the boundary condition (5.2), for the amplitudes of the reflected and refracted waves

$$A_{tg} + A_{tg}^{refl} = A_{tg}^{refr}.$$

In order to find the two quantities A^{refl} and A^{refr} , it is necessary to have a second equation. This is given by the boundary condition for the magnetic field. We restrict ourselves to the case of normal incidence in the xz -plane, which requires no cumbersome calculations. For normal incidence we have

$$E_x = A \exp [i(\omega t - kz)] , \quad E_y = E_z = 0 ;$$

$$H_y = -\epsilon_1^{\frac{1}{2}} A \exp [i(\omega t - kz)] , \quad H_x = H_z = 0 .$$

For the reflected wave we have, analogously,

$$E_x^{refl} = A^{refl} \exp [i(\omega t + kz)] , \quad E_y = E_z = 0 ,$$

$$H_y^{refl} = \epsilon_1^{\frac{1}{2}} A^{refl} \exp [i(\omega t + kz)] , \quad H_z = H_x = 0 .$$

H_y^{refl} differs from H_y in sign, since the reflected wave propagates in the opposite direction and the projection H_y^{refl} is oriented in the positive direction of the y -axis. For the refracted wave

$$E_x^{refr} = A^{refr} \exp [i(\omega t - k^{refr} z)] , \quad E_y = E_z = 0 ,$$

$$H_y^{refr} = -\epsilon_2^{\frac{1}{2}} A^{refr} \exp [i(\omega t - k^{refr} z)] , \quad H_z = H_x = 0 .$$

The conditions of the continuity of the tangential components of the vectors \mathbf{E} and \mathbf{H} for $z = 0$ are written in the form

$$A + A^{refl} = A^{refr} ,$$

$$\epsilon_1^{\frac{1}{2}} (A - A^{refl}) = \epsilon_2^{\frac{1}{2}} A^{refr} .$$

Expressing the quantities A^{refl} and A^{refr} in terms of A , we find

$$A^{\text{refl}} = \frac{1 - (\epsilon_2/\epsilon_1)^{\frac{1}{2}}}{1 + (\epsilon_2/\epsilon_1)^{\frac{1}{2}}} A = \frac{\epsilon_1^{\frac{1}{2}} - \epsilon_2^{\frac{1}{2}}}{\epsilon_1^{\frac{1}{2}} + \epsilon_2^{\frac{1}{2}}} A, \quad (37.5)$$

$$A^{\text{refr}} = \frac{2}{1 + (\epsilon_2/\epsilon_1)^{\frac{1}{2}}} A = \frac{2\epsilon_1^{\frac{1}{2}}}{\epsilon_1^{\frac{1}{2}} + \epsilon_2^{\frac{1}{2}}} A. \quad (37.6)$$

Formulae (37.5) and (37.6) represent a particular case (normal incidence) of the well-known Fresnel formulae derived by Fresnel in 1820 from the general concept of light as a wave motion.

Knowing the amplitudes of the reflected and refracted waves, one can find the time averages of the energy fluxes reflected from the interface and transmitted into the second medium. They are given by the quantities

$$|\sigma^{\text{refl}}| = \frac{c\epsilon_1^{\frac{1}{2}}}{4\pi} \overline{[\mathbf{E}^{\text{refl}} \mathbf{H}^{\text{refl}}]} = \frac{c}{8\pi} \epsilon_1^{\frac{1}{2}} \left(\frac{\epsilon_1^{\frac{1}{2}} - \epsilon_2^{\frac{1}{2}}}{\epsilon_1^{\frac{1}{2}} + \epsilon_2^{\frac{1}{2}}} \right)^2 A^2 = \left(\frac{\epsilon_1^{\frac{1}{2}} - \epsilon_2^{\frac{1}{2}}}{\epsilon_1^{\frac{1}{2}} + \epsilon_2^{\frac{1}{2}}} \right)^2 |\sigma|,$$

$$|\sigma^{\text{refr}}| = \frac{c\epsilon_2^{\frac{1}{2}}}{4\pi} \overline{[\mathbf{E}^{\text{refr}} \mathbf{H}^{\text{refr}}]} = \frac{c}{8\pi} \epsilon_2^{\frac{1}{2}} \left(\frac{2\epsilon_1^{\frac{1}{2}}}{\epsilon_1^{\frac{1}{2}} + \epsilon_2^{\frac{1}{2}}} \right)^2 A^2 = \left(\frac{\epsilon_2}{\epsilon_1} \right)^{\frac{1}{2}} \left(\frac{2\epsilon_1^{\frac{1}{2}}}{\epsilon_1^{\frac{1}{2}} + \epsilon_2^{\frac{1}{2}}} \right)^2 |\sigma|,$$

where σ is the Poynting vector of the incident waves.

The ratio

$$R = \frac{|\sigma^{\text{refl}}|}{|\sigma|} = \left(\frac{\epsilon_1^{\frac{1}{2}} - \epsilon_2^{\frac{1}{2}}}{\epsilon_1^{\frac{1}{2}} + \epsilon_2^{\frac{1}{2}}} \right)^2 = \left(\frac{1 - n_{12}}{1 + n_{12}} \right)^2 \quad (37.7)$$

is called the reflectivity, while the ratio

$$D = \frac{|\sigma^{\text{refr}}|}{|\sigma|} = \left(\frac{\epsilon_2}{\epsilon_1} \right)^{\frac{1}{2}} \left(\frac{2\epsilon_1^{\frac{1}{2}}}{\epsilon_1^{\frac{1}{2}} + \epsilon_2^{\frac{1}{2}}} \right)^2 = 1 - R \quad (37.8)$$

is called the transmissivity.

Formula (37.7) shows that for $\epsilon_2 \approx \epsilon_1$ the effect of reflection of the wave from the interface is small. On the contrary, for $\epsilon_2 \gg \epsilon_1$ or $\epsilon_1 \gg \epsilon_2$ (i.e. for $n_{12} \gg 1$ or $n_{12} \ll 1$), $R \approx 1$ and the wave is almost completely reflected. The considerable reflection of the wave from a dielectric with a large dielectric constant is associated with the fact that in such a dielectric the incident wave excites a large displacement current. Because of this, screening of the external

electric field takes place, and its strength turns out to be close to zero. In the limit, as $\epsilon_2 \rightarrow \infty$, the electric field vector has a node at the interface, while the magnetic field vector has an antinode. However, for any finite value of ϵ_2 the transmissivity D for normal incidence has a small but finite value.

In the case of incidence of the electromagnetic wave at an angle to the normal of the plane of the interface the expressions for the amplitudes (Fresnel formulae) become more complex. However, the general picture is not changed, with the exception of the case $n_{12} < 1$ which calls for special consideration.

If a plane wave falls at an angle θ on the surface of an optically less dense medium, then for a sufficiently small value of n_{12} eq. (37.4) can only be valid for an imaginary value of the angle θ^{refr} . The limiting value of the angle of incidence θ for which θ^{refr} can still have a real value is determined by the condition

$$\frac{\sin \theta_0}{n_{12}} = 1. \quad (37.9)$$

The angle θ_0 is called the critical angle for total reflection. In this case $\theta^{\text{refr}} = \frac{1}{2}\pi$, i.e. the refracted ray slides along the plane of the interface of the media.

For $\theta > \theta_0$ the angle θ^{refr} and, consequently, the value of $\sin \theta^{\text{refr}}$ turn out to be imaginary. In the exponent of the exponential term of the refracted wave the following term arises:

$$i(k^{\text{refr}} \cos \theta^{\text{refr}})z = i \left(\frac{\omega \epsilon_2^{\frac{1}{2}}}{c} (1 - \sin^2 \theta^{\text{refr}})^{\frac{1}{2}} \right) z = \left[- \frac{\omega \epsilon_2^{\frac{1}{2}}}{c} \left(\frac{\sin^2 \theta^{\text{refr}}}{n_{12}^2} - 1 \right)^{\frac{1}{2}} \right] z.$$

This means that in the optically less dense medium damping takes place according to the exponential law

$$|E^{\text{refr}}| \sim \exp \left[- \frac{\omega \epsilon_2^{\frac{1}{2}}}{c} \left(\frac{\sin^2 \theta^{\text{refr}}}{n_{12}^2} - 1 \right)^{\frac{1}{2}} z \right]. \quad (37.10)$$

The effective penetration distance of the electromagnetic field into the optically less dense medium is of the order of magnitude of δ , where

$$\delta \sim \lambda \left(\frac{\sin^2 \theta^{\text{refr}}}{n_{12}^2} - 1 \right)^{-\frac{1}{2}} \quad (37.11)$$

where $\lambda = (\omega \epsilon_2^{1/2} / c)^{-1}$ is the wavelength divided by 2π .

Since there is no absorption in the ideal dielectric considered, the weakening of the field in the optically less dense medium can be associated only with the release of electromagnetic waves backward into the first medium. A direct calculation confirms this. The reflectivity R for $\theta > \theta_0$ turns out to be unity. The effect described is called the phenomenon of total internal reflection.

We cannot, within the framework of this book, dwell on the problem of the appearance of polarization under certain conditions of the reflection of unpolarized waves *.

In conclusion we shall consider briefly the reflection of electromagnetic waves from the surface of conductors. An electromagnetic wave incident on the surface of a conductor induces in it a considerable conduction current. Free charges in the field scatter the wave strongly. Inside the conductor the field is rapidly attenuated. Therefore it is natural to expect the surface of the conductor to possess appreciable reflecting properties. However, part of the energy of the electromagnetic field will be dissipated in the conductor in such a way that the coefficient of reflection will have a value which is somewhat less than unity. If in the formula for R we formally replace the refractive index by its value as given by formula (31.13), then we easily find, taking into account that $\epsilon_1 = 1$ and $\epsilon_2^{1/2} = \epsilon_{\text{met}}^{1/2} = n - i\kappa$; $n \approx \kappa \gg 1$:

$$R = \left| \frac{1-n+i\kappa}{1+n-i\kappa} \right|^2 = \frac{1-2n+n^2+\kappa^2}{1+2n+n^2+\kappa^2} \approx$$

$$\approx \frac{1-2n+2n^2}{1+2n+2n^2} \approx 1 - \frac{2}{n} = 1 - 2 \left(\frac{\omega}{2\pi\sigma} \right)^{\frac{1}{2}}. \quad (37.12)$$

This formula, which is called the Hagen–Rubens formula, is in good agreement with experimental data on the reflection from metal conductors of electromagnetic waves lying in the range of radio and infrared frequencies of the spectrum. For higher frequencies the relations of classical macroscopic electrodynamics turn out to be inapplicable.

The characteristic metallic glitter is due to the high value of the coefficient of reflection in the optical range. In the limiting case of a conductor with an infinitely large conductivity the coefficient of reflection is $R = 1$. This means that a high-frequency electromagnetic field does not penetrate at all

* See, for example, J.A.Stratton, *Electromagnetic theory* (McGraw-Hill, New York, 1941); L.D.Landau and E.M.Lifshitz, *Electrodynamics of continuous media* (Pergamon Press, Oxford, 1960).

into a conductor with an infinitely large conductivity. We shall call such a conductor an ideal conductor. In the range of high frequency fields the ideal conductor is an analogue of the conductor in electrostatics.

Let us write the boundary conditions for the field vectors at the surface of an ideal conductor. Since electric and magnetic fields are absent inside an ideal conductor, it follows from (5.2) – (5.6) that

$$\left. \begin{aligned} E_{tg} &= 0, \\ H_n &= 0, \\ H_{tg} &= \frac{4\pi}{c} i_S, \\ E_n &= \frac{4\pi\omega S}{\epsilon} \end{aligned} \right\} \begin{array}{l} \text{at the surface of} \\ \text{an ideal conductor} \end{array} \quad (37.13)$$

where the vectors **E** and **H** refer to the field in vacuum, and i_S and ω_S are respectively the densities of the surface current and charge. Thus, at the dielectric–ideal conductor interface the electric field vector is directed normally to the surface, whereas the magnetic field vector is directed parallel to the surface. Although in nature there are no ideal conductors, the approximation of an ideal conductor often yields adequately the character of the behaviour of the electromagnetic field at the surface of bodies with a high conductivity.

§38. Wave-guides

The transmission of electromagnetic energy over relatively small distances plays an important role in present day communication engineering. It is brought about by means of the excitation of the electromagnetic field in tubes with metallic walls of various forms and cross-sections, called wave-guides. In this book we cannot dwell on methods for the excitation of fields, and shall confine ourselves only to the study of the process of propagation of electromagnetic waves in wave-guides.

As will be clear from what follows, the propagation of electromagnetic waves in wave-guides differs essentially from the propagation of plane electromagnetic waves which are not limited in space. Therefore the problem of propagation of waves in wave-guides is not only of practical but also of theoretical interest.

To avoid complicated calculations, we shall restrict ourselves to considering a wave-guide of rectangular cross-section with walls made of an ideal conductor. Let the sides of the rectangle be equal to a and b ; for definiteness we assume that $a > b$. We assume that a monochromatic plane electromagnetic wave moves in the wave-guide (in the plane $z = 0$). It is natural to assume that field vectors inside the wave-guide depend on time and the coordinate along the wave-guide according to the law

$$\mathbf{E}, \mathbf{H} \sim \exp [i(\omega t - k_z z)] . \quad (38.1)$$

Substituting (38.1) into the wave equations, we find

$$\frac{\partial^2 \mathbf{E}}{\partial x^2} + \frac{\partial^2 \mathbf{E}}{\partial y^2} = \left(k_z^2 - \frac{\omega^2}{v^2} \right) \mathbf{E} , \quad (38.2)$$

$$\frac{\partial^2 \mathbf{H}}{\partial x^2} + \frac{\partial^2 \mathbf{H}}{\partial y^2} = \left(k_z^2 - \frac{\omega^2}{v^2} \right) \mathbf{H} . \quad (38.3)$$

Here and below \mathbf{E} and \mathbf{H} are unknown functions of the coordinates x and y , since their dependence on z is already taken into account. The connection between the vectors \mathbf{E} and \mathbf{H} is determined by the equations for $\nabla \times \mathbf{E}$ and $\nabla \times \mathbf{H}$. Substituting (38.1) into these equations, we find

$$\frac{\partial E_z}{\partial y} + ik_z E_y = -\frac{i\mu\omega}{c} H_x , \quad (38.4)$$

$$-ik_z E_x - \frac{\partial E_z}{\partial x} = -\frac{i\mu\omega}{c} H_y , \quad (38.5)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\frac{i\mu\omega}{c} H_z , \quad (38.6)$$

$$\frac{\partial H_z}{\partial y} + ik_z H_y = \frac{i\epsilon\omega}{c} E_x , \quad (38.7)$$

$$-ik_z H_x - \frac{\partial H_z}{\partial x} = \frac{i\epsilon\omega}{c} E_y , \quad (38.8)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = \frac{i\epsilon\omega}{c} E_z . \quad (38.9)$$

Formulae (38.4)–(38.9) allow the vector components E_x, E_y, H_x and H_y to be expressed in terms of E_z and H_z . A simple calculation gives

$$E_x = \frac{1}{c^2 k_z^2 - \epsilon \mu \omega^2} \left(i c^2 k_z \frac{\partial E_z}{\partial x} + i \omega \mu c \frac{\partial H_z}{\partial y} \right), \quad (38.10)$$

$$E_y = \frac{1}{-\omega^2 \epsilon \mu + c^2 k_z^2} \left(i c^2 k_z \frac{\partial E_z}{\partial y} - i \omega \mu c \frac{\partial H_z}{\partial x} \right), \quad (38.11)$$

$$H_x = \frac{1}{-\epsilon \mu \omega^2 + c^2 k_z^2} \left(-i \epsilon \omega c \frac{\partial E_z}{\partial y} + i c^2 k_z \frac{\partial H_z}{\partial x} \right), \quad (38.12)$$

$$H_y = \frac{1}{c^2 k_z^2 - \epsilon \mu \omega^2} \left(i \epsilon \omega c \frac{\partial E_z}{\partial x} + i c^2 k_z \frac{\partial H_z}{\partial y} \right). \quad (38.13)$$

At the surface of the wave-guide the boundary conditions for an ideal conductor (see §37) must be fulfilled.

We first try to find the solution of eqs. (38.2) and (38.3) in the form of transverse plane waves, i.e. we assume

$$E_z = H_z = 0.$$

Then from formulae (38.10)–(38.13) it is clear that all components of the field are equal to zero, provided $c^2 k_z^2 - \epsilon \mu \omega^2 \neq 0$. If, on the contrary, $c^2 k_z^2 = \epsilon \mu \omega^2$, as is the case of a monochromatic plane wave in a non-limited medium propagating with a velocity $c/(\epsilon \mu)^{1/2}$, then (38.2) gives

$$\frac{\partial^2 H(x, y)}{\partial x^2} + \frac{\partial^2 H(x, y)}{\partial y^2} = 0,$$

i.e. the magnetic field satisfies the two-dimensional Laplace equation, and over the entire closed boundary of the region (at the edges of the rectangle $x = 0, a; y = 0, b$) the magnetic field by virtue of (37.13) is directed along a tangent to the boundary. In the mathematical theory it is shown* that the only solution of such a boundary value problem is $H = 0$. If there is no

* See, for example, V.I. Smirnov, *A course of higher mathematics* (Pergamon Press, Oxford, 1964).

magnetic field in the wave, then, obviously, the electric field is also equal to zero.

Thus we see that transverse electromagnetic waves cannot propagate in a rectangular wave-guide with ideally conducting walls. It should be stressed that this conclusion is not associated with the form of the wave-guide, but refers to any wave-guide made in the form of a simple tube of any cross-section whose walls are ideally conducting. In the case where the boundaries of the region are open, for example, in the case of the gap between two infinite ideally conducting plane surfaces, or in the case of multiply connected space, for example a wave-guide formed by two concentric cylinders, this conclusion is no longer valid. In such systems the propagation of transverse electromagnetic waves is in principle possible.

The meaning of this result is very simple. In a transverse wave in a wave-guide the lines of the magnetic field must be directed along a tangent to the walls and must have the character of closed curves. In this case the lines of the field do not enter into the ideal conductor and therefore do not enclose any tubes of conduction current. A longitudinal component of displacement current is absent, so that the lines of the magnetic field do not enclose any tubes of displacement current either. However, from general considerations it is clear that the lines of a magnetic field which do not enclose any current tubes cannot exist.

This kind of reasoning also allows one to understand why the existence of transverse waves is possible in the open wave-guide or in a wave-guide formed by concentric cylinders. In an unbounded space the lines of the magnetic field can be open and go off to infinity. In a wave-guide formed by concentric cylinders they can enclose the tubes of current flowing on the surface of the internal cylinder. In both cases the existence of magnetic and electric fields travelling along the wave-guide in the form of transverse waves is possible.

It turns out, however, that the absence of transverse waves in a wave-guide does not mean that the propagation of an electromagnetic field in it is impossible. As will now be shown, the formation of longitudinal waves, which cannot exist in unbounded space, is possible in a wave-guide. By longitudinal waves we understand waves which have a non-zero field component in the direction of propagation of the wave. From formulae (38.10) – (38.13) it is clear that two independent possibilities must be considered:

$$1) \quad E_z \neq 0, \quad H_z = 0,$$

$$2) \quad H_z \neq 0, \quad E_z = 0.$$

In the first case the magnetic field in the wave has two components, H_x and H_y , and is purely transverse. Such waves are usually called transverse magnetic waves or, briefly, TM-waves. The electric field has a longitudinal component and two transverse components.

In the second case the electric field of the wave has a transverse character and the wave is called correspondingly a transverse electric wave or, briefly, TE-wave.

In the case of TM-waves, on the basis of (38.2) we have

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} = \left(k_z^2 - \frac{\omega^2}{v^2} \right) E_z \quad (38.14)$$

and

$$E_z = 0 \quad \text{for} \quad x = 0, a; \quad y = 0, b, \quad (38.15)$$

by virtue of (37.13). Below we shall show that it is sufficient to satisfy only these boundary conditions. In order to satisfy eq. (38.14) and the boundary conditions (38.15), one must write for E_z

$$E_z = A \sin(k_x x) \sin(k_y y) \exp[i(\omega t - k_z z)] , \quad (38.16)$$

where

$$k_x^2 + k_y^2 + k_z^2 = \omega^2/v^2 \quad (38.17)$$

and, moreover,

$$k_x = \pi m/a; \quad k_y = \pi n/b, \quad (38.18)$$

where m and n are integers which are not equal to zero. This solution has the form of waves travelling in the positive direction of the z -axis and standing waves in the planes $z = \text{const}$.

The transverse magnetic wave corresponding to the numbers m and n is usually denoted by TM_{mn} . If the value of E_z is known, then from (38.10) – (38.13) one can easily find the remaining components of the field. We shall not write down the corresponding formulae, and shall only point out the following fact. It is easily understood that the boundary conditions for all components of the electric and magnetic fields are automatically fulfilled. For example, in the plane $y = 0$ the normal component of the magnetic field is

$$H_n = H_y \sim \frac{\partial E_z}{\partial x} \sim \sin \frac{\pi n}{b} y \Big|_{y=0} = 0 ;$$

analogously,

$$E_{tg} = E_x = E_z = 0 .$$

Hence, as we have already pointed out, it is sufficient to satisfy the boundary conditions (38.15). Thus, at the walls of the wave-guide the lines of the magnetic field are tangential to the surface. They form closed curves embracing the longitudinal lines of the electric field.

Let us discuss formula (38.17) in more detail. We rewrite it in the form

$$k_z = \left(\frac{\omega^2}{v^2} - \frac{\pi^2 m^2}{a^2} - \frac{\pi^2 n^2}{b^2} \right)^{\frac{1}{2}} . \quad (38.19)$$

For given values of m and n the quantity k_z has a real value only for

$$\frac{1}{\lambda^2} = \frac{\omega^2}{4\pi^2 v^2} \geq \frac{(\omega_{crit}^{m,n})^2}{4\pi^2 v^2} = \left(\frac{\pi^2 m^2}{a^2} + \frac{\pi^2 n^2}{b^2} \right) \frac{1}{4\pi^2} = \frac{1}{\lambda_{crit}^2} , \quad (38.20)$$

where $\omega_{crit}^{m,n}$ is a certain critical frequency, and λ_{crit} is the corresponding wavelength.

If k_z turns out to be imaginary, which corresponds to $\omega < \omega_{crit}$, then we arrive at an exponentially decreasing expression for E_z and the other field components instead of a wave travelling along the z -axis. Thus, only waves with $\omega > \omega_{crit}^{m,n}$ or $\lambda < \lambda_{crit}$ can propagate through the wave-guide.

The largest value of the wavelength of the propagating wave, λ_{crit}^{max} , is obtained for the TM_{11} -wave ($m = 1$ and $n = 1$) which is the wave of TM-type of lowest order. Namely,

$$\lambda_{TM_{11}} = \frac{2ab}{(a^2 + b^2)^{\frac{1}{2}}} .$$

The phase velocity of the TM_{mn} -wave is obviously equal to

$$\begin{aligned} v_{phase} &= \frac{\omega}{k_z} = \frac{c}{(\epsilon\mu)^{\frac{1}{2}} (1 - \lambda^2 [(m/2a)^2 + (n/2b)^2])^{\frac{1}{2}}} = \\ &= \frac{c}{(\epsilon\mu)^{\frac{1}{2}} [1 - (\lambda^2/\lambda_{crit}^2)]^{\frac{1}{2}}} = \frac{v}{[1 - (\lambda^2/\lambda_{crit}^2)]^{\frac{1}{2}}} . \end{aligned} \quad (38.21)$$

Since $\lambda < \lambda_{\text{crit}}$ always, the phase velocity of waves in a wave-guide is always larger than the velocity of light v in a corresponding medium, and for $\lambda \rightarrow \lambda_{\text{crit}}$ it increases indefinitely. In particular, if no dielectric is introduced inside the wave-guide (i.e. $\epsilon = \mu = 1$), v_{phase} is always larger than the velocity of light in vacuum c .

The group velocity can easily be found from formula (38.17):

$$v_{\text{group}} = \frac{d\omega}{dk_z} = \frac{c}{(\epsilon\mu)^{\frac{1}{2}}} \left(1 - \frac{\lambda^2}{\lambda_{\text{crit}}^2} \right)^{\frac{1}{2}}. \quad (38.22)$$

The group velocity is always smaller than the velocity $v = c/(\epsilon\mu)^{\frac{1}{2}}$ and tends to zero as λ approaches λ_{crit} .

From (38.21) and (38.22) it is clear that v_{phase} and v_{group} satisfy the general relation $v_{\text{phase}} v_{\text{group}} = v^2$.

Let us now consider TE-waves. For such waves we have

$$\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} = \left(k_z^2 - \frac{\omega^2}{v^2} \right) H_z. \quad (38.23)$$

By means of (38.10) and (38.11) the boundary conditions for the tangential component of the electric field can be expressed in terms of the derivatives of H_z . Then instead of

$$E_x = 0 \quad \text{for } y = 0, b$$

we have

$$\partial H_z / \partial y = 0 \quad \text{for } y = 0, b. \quad (38.24)$$

Analogously, instead of

$$E_y = 0 \quad \text{for } x = 0, a$$

we have

$$\partial H_z / \partial x = 0 \quad \text{for } x = 0, a. \quad (38.25)$$

The solution of eq. (38.23) for the boundary conditions (38.24) and (38.25) can be written in the form of a TE-wave:

$$H_z = B \cos(k_x x) \cos(k_y y) \exp [i(\omega t - k_z z)] , \quad (38.26)$$

where

$$k_x = \pi m/a ; \quad k_y = \pi n/b .$$

Here m and n are integers. In contrast to the case of TM-waves, each of these numbers (but not both of them simultaneously) can take on the value zero. Formulae (38.19) – (38.22) remain valid for TE-waves. In them, however, one of the numbers m or n can be set equal to zero. The TE_{10} -wave is the TE-wave of lowest order. For the TE_{10} -wave the critical wavelength is

$$\lambda_{\text{crit}}^{\text{max}} = 2a .$$

The most important conclusion which can be drawn from the theory described is that the transverse character of plane waves is closely associated with the unlimited dimensions of the medium in which they are propagating. When waves are propagating in a limited region, purely transverse waves can exist only under particular conditions (multiply connected regions). Under usual conditions the waves do have a longitudinal component of the electric or magnetic field. We recall also, that in an unlimited space, but near the emitter, the electromagnetic wave has non-zero longitudinal (radial) field components E_r and H_r (see §25 and §26 of Part I).

In the calculations for a wave-guide of the simplest form which we have presented above we have not taken into account effects which in practice play an important part. Such an effect is, first of all, the power loss associated with the non-ideal character of the walls of the wave-guide and the dielectric which fills it. Also we have not considered the theory of wave-guides of a more complex form and wave-guides which are inhomogeneous over their length.

Neither have we touched upon numerous problems associated with various applications of wave-guides, including their use in linear accelerators for charged particles. For all these problems we refer the reader to the comprehensive specific literature *.

* See, for example, L.A.Vainshtein, *Elektromagnitnye volny (Electromagnetic waves)*, Sov. Radio, 1957; S.Ramo, J.R.Whinnery and Th.Van Duzer, *Fields and waves in communication electronics* (Wiley, New York, 1965).

§39. The passage of fast particles through matter

The problem of the energy loss and radiation by fast particles moving in matter is of great importance in contemporary physics.

As we shall see below, there are several different mechanisms of energy loss by particles moving in matter.

First of all it should be recalled that, as we have seen in §43 of Part I and §26 of Part II, particles which undergo collisions with atoms and are deflected by the nuclear field emit bremsstrahlung. According to eq. (26.15) of Part II the bremsstrahlung of ultra-relativistic particles is inversely proportional to the mass of the particle and plays an important part for light particles (electrons). When heavy charged particles (protons, ions) pass through matter, other sources of energy loss usually play a greater part. The charged particle moving in matter interacts with the atoms of the substance and polarizes them. In other words, the moving particle produces a field in the substance. This field induces a reaction on the particle itself which decelerates it. The energy loss of the particle associated with this deceleration is obviously equal to the work done by the decelerating force. This is called the polarization energy loss, since the decelerating field is the polarization field produced by the moving particle. The medium which is polarized by the particle can however emit transverse electromagnetic waves. To do this it is necessary that the polarization caused by the particle moving in the medium does not follow the particle. The moving particle leaves the medium in a polarized state and the medium emits the excess energy in the form of transverse electromagnetic waves. This source of energy loss is the Cerenkov-Vavilov radiation or briefly the Cerenkov radiation.

We stress that Cerenkov-Vavilov radiation is not associated with any acceleration, since it is radiation emitted by the medium and not by the moving charge. On the other hand, as will be shown in the following, Cerenkov radiation is possible only in the case of motion of a particle with a velocity v which is larger than the velocity of propagation of the field in the medium c/n . This restriction expresses the above-mentioned condition of the lag of the polarization field behind the moving particle. We now go on to the simultaneous calculation of the polarization losses and Cerenkov losses of one particle moving in a macroscopically homogeneous and isotropic medium. Our purpose is to find the field produced in the medium by a single charge moving with a velocity v .

We disregard the decrease in velocity during the motion and consider the velocity to be constant in magnitude and direction. We assume the medium to be transparent, so that the imaginary part of the dielectric constant is $\epsilon^i \rightarrow 0$. The magnetic permeability μ is assumed to be equal to unity.

The current density in the medium corresponding to a uniform motion of one particle can be written in the form

$$\mathbf{j} = ev\delta(\mathbf{r}-\mathbf{v}t). \quad (39.1)$$

Hence Maxwell's equations will have the form

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad (39.2)$$

$$\nabla \times \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} + \frac{4\pi e}{c} \mathbf{v}\delta(\mathbf{r}-\mathbf{v}t). \quad (39.3)$$

We shall discuss below the conditions under which one can assume a medium to be continuous and disregard its atomic structure describing it by macroscopic quantities.

We seek the solution of Maxwell's equations by means of expansion in a Fourier integral. Writing formulae (33.1) – (33.4) for the field vectors and for the vector \mathbf{j} the expression

$$\mathbf{j}(\mathbf{r}, t) = \frac{1}{(2\pi)^4} \int \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)] \mathbf{j}(\mathbf{k}, \omega) d\mathbf{k} d\omega, \quad (39.4)$$

we write the system (39.2) – (39.3) in the form

$$i[\mathbf{k} \times \mathbf{E}(\mathbf{k}, \omega)] = i \frac{\omega}{c} \mathbf{B}(\mathbf{k}, \omega), \quad (39.5)$$

$$i[\mathbf{k} \times \mathbf{H}(\mathbf{k}, \omega)] = -i \frac{\omega}{c} \mathbf{D}(\mathbf{k}, \omega) + \frac{4\pi}{c} \mathbf{j}(\mathbf{k}, \omega). \quad (39.6)$$

Forming the vector product of (39.5) and \mathbf{k} , thereupon eliminating $\mathbf{k} \times \mathbf{H}$ from (39.6), and making use of the constitutive equation (33.4), we arrive at the equation

$$\left\{ k^2 \delta_{ij} - k_i k_j - \frac{\omega^2}{c^2} \epsilon_{ij} \right\} E_j = i \frac{4\pi\omega}{c^2} j_i, \quad (39.7)$$

which differs from (33.15) only in the presence of the right-hand side. The quantity j_i in (39.7) denotes the Fourier component of the current density

$\mathbf{j}(\mathbf{k}, \omega)$. For the isotropic medium we substitute for ϵ_{ij} its value from formula (33.9). We then find

$$\{k^2 \delta_{ij} - k_i k_j\} E_j - \frac{\omega^2 \epsilon_{\perp}}{c^2} \left\{ \delta_{ij} - \frac{k_i k_j}{k^2} \right\} E_j - \frac{\omega^2 \epsilon_{\parallel}}{c^2} \frac{k_i k_j}{k^2} E_j = i \frac{4\pi\omega}{c^2} j_i,$$

or, in the vector form

$$\left(k^2 - \frac{\omega^2 \epsilon_{\perp}}{c^2} \right) \left(\mathbf{E} - \frac{\mathbf{k}(\mathbf{k} \cdot \mathbf{E})}{k^2} \right) - \frac{\omega^2 \epsilon_{\parallel}}{c^2} \frac{\mathbf{k}(\mathbf{k} \cdot \mathbf{E})}{k^2} = i \frac{4\pi\omega}{c^2} \mathbf{j}. \quad (39.8)$$

Forming the scalar product of (39.8) and the vector \mathbf{k} , we obtain

$$\mathbf{k} \cdot \mathbf{E} = -i \frac{4\pi}{\omega \epsilon_{\parallel}} (\mathbf{k} \cdot \mathbf{j}). \quad (39.9)$$

Substituting (39.9) into (39.8), we find

$$\left(k^2 - \frac{\omega^2 \epsilon_{\perp}}{c^2} \right) \mathbf{E} = -i \frac{4\pi\omega}{k^2} \left\{ \left(k^2 - \frac{\omega^2 \epsilon_{\perp}}{c^2} \right) \frac{(\mathbf{k} \cdot \mathbf{j})\mathbf{k}}{\omega^2 \epsilon_{\parallel}} + \frac{\mathbf{k}(\mathbf{k} \cdot \mathbf{j})}{c^2} - \frac{k^2}{c^2} \mathbf{j} \right\}.$$

Hence

$$\mathbf{E}(\omega, \mathbf{k}) = -\frac{4\pi\omega i}{k^2} \left\{ \frac{(\mathbf{k} \cdot \mathbf{j}(\omega, \mathbf{k}))\mathbf{k}}{\omega^2 \epsilon_{\parallel}} + \frac{\mathbf{k}(\mathbf{k} \cdot \mathbf{j}(\omega, \mathbf{k})) - k^2 \mathbf{j}(\omega, \mathbf{k})}{c^2(k^2 - \omega^2 \epsilon_{\perp})} \right\}. \quad (39.10)$$

Going from the Fourier component to the electric field, we have

$$\mathbf{E}(\mathbf{r}, t) = \int \mathbf{E}(\omega, \mathbf{k}) \exp [i(\mathbf{k} \cdot \mathbf{r} - \omega t)] d\mathbf{k} d\omega = \mathbf{E}_1(\mathbf{r}, t) + \mathbf{E}_2(\mathbf{r}, t). \quad (39.11)$$

The first term $\mathbf{E}_1(\mathbf{r}, t)$ has the form

$$\mathbf{E}_1 = -4i\pi \int \frac{d\mathbf{k} d\omega}{k^2} \frac{(\mathbf{k} \cdot \mathbf{j})\mathbf{k}}{\omega \epsilon_{\parallel}}. \quad (39.12)$$

We now substitute the value of $\mathbf{j}(\omega, \mathbf{k})$, making use of formula (III.8'). We have, obviously,

$$\mathbf{j}(\omega, \mathbf{k}) = \frac{e\mathbf{v}}{(2\pi)^3} \delta(\omega - \mathbf{k} \cdot \mathbf{v}). \quad (39.13)$$

Then E_1 assumes the form

$$\begin{aligned} E_1 &= -i \frac{4\pi e}{(2\pi)^3} \int d\omega \frac{dk}{k^2} \frac{k(k \cdot v)}{\omega \epsilon_{||}} \delta(\omega - k \cdot v) \exp [i(k \cdot r - \omega t)] = \\ &= -\frac{ie}{2\pi^2} \int dk \exp [i(k \cdot r - k \cdot vt)] \frac{k}{k^2 \epsilon_{||}(k \cdot v, k)}, \end{aligned} \quad (39.14)$$

where $\epsilon_{||}(k \cdot v, k)$ is the value of $\epsilon_{||}(\omega, k)$ for $\omega = k \cdot v$.

The field E_1 produced by the moving charged particle in the medium induces a reaction on it. The particle is acted upon by a force

$$F = e(E_1)_{r=vt} + \frac{e}{c} (v \times H)_{r=vt},$$

where the index shows that the value of the field at each instant of time must be taken at the point $r = vt$ at which the particle is found. The energy loss of the particle (we shall relate it to unit path in matter) is equal to the work done by this force per unit path. Since the magnetic part of the Lorentz force $ec^{-1}(v \times H)$ does no work the work done by the force F per unit path is equal to

$$\begin{aligned} W_1 &= e \left(\frac{v}{v} \cdot E_1 \right) = -\Delta E_1 = \\ &= \text{Re} \left\{ -\frac{ie^2}{2\pi^2 v} \int dk \exp [i(k \cdot r - k \cdot vt)] \frac{k \cdot v}{k^2} \frac{1}{\epsilon_{||}(k \cdot v, k)} \right\}_{r=vt} = \\ &= \text{Re} \left\{ -\frac{ie^2}{4\pi^2 v} \iint dk d\omega \frac{\omega \delta(\omega - k \cdot v)}{k^2 \epsilon_{||}(\omega, k)} \right\}. \end{aligned} \quad (39.15)$$

The notation Re is introduced in order to show that the expression for the work W_1 is real. This work has a clear physical meaning. It is determined by the value of $\epsilon_{||}$ and represents the work done on the particle by the longitudinal polarization produced by the particle. Thus, dW_1/dr is a measure of the energy loss ΔE_1 of the particle per unit path. This energy goes into the production of the longitudinal polarization. Later, in calculating ΔE_1 , we shall discuss the question as to what processes actually occur in a medium when a moving particles produces a longitudinal polarization in it.

We now come back to the second term of (39.11). According to formulae (39.11) and (39.8), we have

$$\begin{aligned}
 E_2(\mathbf{r}, t) &= i \frac{4\pi e}{(2\pi)^3} \int d\mathbf{k} d\omega \exp [i(\mathbf{k} \cdot \mathbf{r} - \omega t)] \frac{k^2 \mathbf{v} - \mathbf{k}(\mathbf{k} \cdot \mathbf{v})}{c^2(k^2 - \omega^2 c^{-2} \epsilon_{\perp})} \delta(\omega - \mathbf{k} \cdot \mathbf{v}) \frac{\omega}{k^2} = \\
 &= \frac{ie}{2\pi^2 c^2} \int d\mathbf{k} \exp [i(\mathbf{k} \cdot \mathbf{r} - \mathbf{k} \cdot \mathbf{v} t)] \frac{(\mathbf{k} \cdot \mathbf{v})}{k^2} \frac{k^2 \mathbf{v} - \mathbf{k}(\mathbf{k} \cdot \mathbf{v})}{k^2 - (\mathbf{k} \cdot \mathbf{v})^2 c^{-2} \epsilon_{\perp}(\mathbf{k} \cdot \mathbf{v}, \mathbf{k})}.
 \end{aligned} \quad (39.16)$$

Correspondingly, the work done on the charge due to the field E_2 and the energy loss ΔE_2 of the particle per unit path are equal to

$$\begin{aligned}
 W_2 = -\Delta E_2 &= \operatorname{Re} \left\{ \frac{ie^2}{2\pi^2 v c^2} \int d\mathbf{k} \frac{(\mathbf{k} \cdot \mathbf{v}) [v^2 - (\mathbf{k} \cdot \mathbf{v})^2 k^{-2}]}{k^2 - (\mathbf{k} \cdot \mathbf{v})^2 c^{-2} \epsilon_{\perp}(\mathbf{k} \cdot \mathbf{v}, \mathbf{k})} \right\} \bigg|_{\mathbf{r} = \mathbf{v} t} = \\
 &= \operatorname{Re} \left\{ \frac{ie^2}{2\pi^2 v c^2} \int d\mathbf{k} d\omega \frac{\omega (v^2 - \omega^2 k^{-2}) \delta(\omega - \mathbf{k} \cdot \mathbf{v})}{k^2 - \omega^2 c^{-2} \epsilon_{\perp}(\omega, \mathbf{k})} \right\}.
 \end{aligned} \quad (39.17)$$

The energy loss ΔE_2 determines the Cerenkov loss of the particle. This is associated with the excitation of a transverse electromagnetic field in the medium, i.e. with the radiation of the polarized medium.

Taking account of spatial dispersion allows us to separate clearly the two types of energy loss: polarization loss and Cerenkov loss.

In what follows we shall restrict ourselves to the calculation of losses in a medium without spatial dispersion*. For this, according to (33.11), we have to set in the formulae obtained

$$\epsilon_{\parallel}(\omega, \mathbf{k}) = \epsilon_{\perp}(\omega, \mathbf{k}) = \epsilon(\omega) \big|_{\omega = \mathbf{k} \cdot \mathbf{v}}. \quad (39.18)$$

We pass on to the calculation of the integrals in formulae (39.15) and (39.17). We have, taking into account (39.18),

$$\Delta E_1 = -\operatorname{Re} \left\{ \frac{ie^2}{2\pi^2 v} \int d\mathbf{k} \frac{\omega}{k^2 \epsilon(\omega)} \right\}. \quad (39.19)$$

* For losses taking into account spatial dispersion see, for example, V.P.Silin and A.A.Rukhadze, *Elektromagnitnye svoistva plazmy i plazmopodobnykh sred* (Electromagnetic properties of plasma and plasma-like media), Moscow, 1961; V.D.Shafranov, *Elektromagnitnye volny v plazme* (Electromagnetic waves in plasma), collected papers *Voprosy fiziki plazmy* (Problems of plasma physics), Vol. II, Moscow, 1963. V.L.Ginzburg, *Propagation of electromagnetic waves in plasma* (North-Holland Publ. Co., Amsterdam, 1961).

We have already pointed out above that all our considerations will refer to transparent media, more precisely to media with a negligibly small absorption. This means that in the expression for the dielectric constant $\epsilon(\omega) = \epsilon^r(\omega) + i\epsilon^i(\omega)$ we have to go to the limit $\epsilon^i(\omega) \rightarrow 0$. This allows us to greatly simplify all subsequent calculations. That is, we can write

$$\operatorname{Re} \frac{i}{\epsilon(\omega)} = \operatorname{Re} \frac{i}{\epsilon^r(\omega) + i\epsilon^i(\omega)} = \operatorname{Re} \frac{i(\epsilon^r + i\epsilon^i)}{(\epsilon^r)^2 + (\epsilon^i)^2} = \frac{(-\epsilon^i)}{(\epsilon^r)^2 + (\epsilon^i)^2}.$$

In the limit $\epsilon^i \rightarrow 0$, we can make use of formula (III.4'), Vol. 1, and write

$$\lim_{\epsilon^i \rightarrow 0} \operatorname{Re} \frac{i}{\epsilon(\omega)} = - \lim_{\epsilon^i \rightarrow 0} \left[\frac{i}{(\epsilon^r)^2 + (\epsilon^i)^2} \right] = - \pi \delta[\epsilon(\omega)]. \quad (39.20)$$

Then the formula for polarization loss assumes the form

$$\Delta E_1 = \frac{e^2}{2\pi v} \int \frac{\omega}{k^2} \delta[\epsilon(\omega)] dk = \frac{e^2}{2\pi v} \int \frac{(\mathbf{k} \cdot \mathbf{v})}{k^2} \delta[\epsilon(\mathbf{k} \cdot \mathbf{v})] dk, \quad (39.21)$$

where we have made use of the fact that only the frequency $\omega = \mathbf{k} \cdot \mathbf{v}$ gives a contribution to the loss. If the explicit form of $\epsilon(\omega)$ is known, then the integral in (39.21) can easily be calculated. In §46 such a calculation will be carried out for the case of a plasma.

We note that formula (39.21) does not contain the mass of the particle. This is quite natural, since the energy losses are associated with the polarization of the medium. Polarization losses are therefore the main source of energy loss for heavy particles moving in matter for a wide energy range. Light particles, for example electrons of relatively high energies, lose most of their energy due to bremsstrahlung.

We now go on to the calculation of the Cerenkov losses, which can be carried out by the same method. We have

$$\Delta E_2 = - \frac{e^2}{2\pi^2 v c^2} \operatorname{Re} \left\{ i \int dk \frac{v^2 - (\mathbf{k} \cdot \mathbf{v})^2 k^{-2}}{k^2 - (\mathbf{k} \cdot \mathbf{v})^2 c^{-2} \epsilon(\omega)} \right\}. \quad (39.22)$$

We choose the direction of motion of the particles as the z -axis and introduce new variables

$$\omega = \mathbf{k} \cdot \mathbf{v} = k_z v,$$

$$q^2 = k_x^2 + k_y^2 = k^2 - \omega^2 v^{-2}.$$

Then

$$dk = dk_x dk_y dk_z = q dq d\varphi \frac{d\omega}{v}.$$

In the new variables we have

$$\begin{aligned} \Delta E_2 &= -\frac{e^2}{2\pi^2 v^2 c^2} \operatorname{Re} \left\{ i \int_{-\infty}^{\infty} \omega d\omega \int_0^{\infty} \frac{q dq}{q^2 + \omega^2 v^{-2}} \right. \\ &\quad \left. \int_0^{2\pi} d\varphi \frac{q^2 v^2}{q^2 + \omega^2 [v^{-2} - \epsilon(\omega) c^{-2}]} \right\} = \\ &= -\frac{e^2}{\pi c^2} \operatorname{Re} \left\{ i \int_{-\infty}^{\infty} \omega d\omega \int_0^{\infty} \frac{q^3 dq}{q^2 + \omega^2 c^{-2}} \frac{1}{q^2 + \omega^2 v^{-2} - \omega^2 c^{-2} \epsilon(\omega)} \right\}. \end{aligned}$$

For a non-absorbing medium we can write, analogously to (39.20),

$$\lim_{\epsilon \rightarrow 0} \operatorname{Re} \left\{ \frac{i}{q^2 + \omega^2 v^{-2} - c^{-2} \epsilon(\omega)} \right\} = -\pi \delta \left[q^2 + \omega^2 \left(\frac{1}{v^2} - \frac{\epsilon(\omega)}{c^2} \right) \right].$$

Hence

$$\Delta E_2 = \frac{e^2}{\pi c^2} \int_{-\infty}^{\infty} \omega d\omega \int_0^{\infty} \frac{q^3 dq}{q^2 + \omega^2 v^{-2}} \delta \left[q^2 + \omega^2 \left(\frac{1}{v^2} - \frac{\epsilon(\omega)}{c^2} \right) \right].$$

Introducing the new variable $u = q^2$, we have

$$\begin{aligned} &\int_0^{\infty} \frac{q^3 dq}{q^2 + \omega^2 v^{-2}} \delta \left[q^2 + \omega^2 \left(\frac{1}{v^2} - \frac{\epsilon(\omega)}{c^2} \right) \right] = \\ &= \frac{1}{2} \int_0^{\infty} \frac{u du}{u + \omega^2 v^{-2}} \delta \left[u - \left(\frac{\epsilon(\omega)}{c^2} \omega^2 - \frac{\omega^2}{v^2} \right) \right] = \\ &= \begin{cases} \frac{1}{2} \left(1 - \frac{c^2}{v^2 \epsilon(\omega)} \right), & v > c/n \\ 0, & v < c/n, \end{cases} \end{aligned}$$

where $n = \epsilon^{\frac{1}{2}}(\omega)$.

Hence we arrive finally at the following expression for the energy loss per unit path due to the Cerenkov–Vavilov transverse electromagnetic radiation:

$$\Delta E_2 = \frac{e^2}{2c^2} \int_{-\infty}^{\infty} \omega \, d\omega \left(1 - \frac{c^2}{v^2 \epsilon(\omega)} \right) = \frac{e^2}{c^2} \int_0^{\infty} \left(1 - \frac{c^2}{v^2 \epsilon(\omega)} \right) \omega \, d\omega, \quad (39.23)$$

since $\epsilon(\omega)$ is an even function of its argument (see (32.9)).

Thus we obtain the following basic properties of the Cerenkov–Vavilov radiation:

- (1) it arises only for particles moving with a velocity which is larger than $v_{\text{cut-off}} = c/n$ (cut-off of radiation),
- (2) it depends on the charge of the particles, but not on their mass,
- (3) the radiation lies in the visible spectrum and in part of the ultraviolet spectrum. For shorter waves $n < 1$ (see §34) and the radiation is no longer possible,
- (4) the energy emitted per unit path in unit frequency interval has a characteristic spectral distribution (39.23),
- (5) the radiation arising at a given point of the trajectory propagates over the surface of a cone with the vertex at this point and the axis coinciding with the direction of motion of the particle. The angle θ of the cone is determined by the condition

$$\cos \theta = \frac{c}{nv}. \quad (39.24)$$

In conclusion we stress that the energy loss due to Cerenkov–Vavilov radiation is very small and amounts to only about 0.1% of the energy loss of very fast particles in matter. The rest is determined by bremsstrahlung and a number of other phenomena. The importance of Cerenkov radiation lies in the fact that new detectors of fast particles are based on the use of it. These detectors, which are called Cerenkov counters, have now become one of the chief working instruments in the field of high-energy particle physics. By observing the Cerenkov radiation one can not only record the passage of particles but also according to (39.24) determine directly the magnitude and direction of their velocity, according to (39.23) determine their charge, and separate particles with the same momentum but a different mass (making use of the existence of the cut-off $v_{\text{cut-off}} = p_{\text{cut-off}}/m$), and so on.

In conclusion we shall discuss the problem of the limits of applicability of the macroscopic treatment. Fast particles actually interact with individual

atoms or, more precisely, with atomic electrons. For the macroscopic description of the process to be valid it is necessary that during the time of flight of the particle past the atom the intra-atomic electrons do not undergo an appreciable displacement. The particle in flight is moving in the quasi-stationary field of many atoms.

Thus the velocity of the particle must be large compared with the characteristic velocity of atomic electrons.

An interesting source of energy loss due to radiation by a uniformly moving particle is the so-called transitional radiation. It arises when a particle passes from one medium to another. In order to simplify the formulae we assume that a particle which is moving uniformly in vacuum with a velocity $v \sim c$ falls on the boundary of a medium with a dielectric constant $\epsilon(\omega)$. We assume the boundary of the medium to be the plane $z = 0$, and the incident direction of the particle to be normal to the boundary. We choose the instant of impact of the particle on the boundary to be the time $t = 0$, and find the vector potential of the field of the particle according to formula (32.10) of Part I.

Obviously, for the Fourier component A_ω at a large distance from the charge we have

$$\begin{aligned} A_\omega &= \frac{e}{2\pi cr} e^{ikr} \int_{-\infty}^{\infty} v(t) \exp [i(k \cdot r(t) - \omega t)] dt = \\ &= \frac{ev}{2\pi cr} e^{ikr} \left\{ \int_{-\infty}^0 \exp [it(k \cdot v \epsilon^{\frac{1}{2}} - \omega)] dt + \int_0^{\infty} \exp [it(k \cdot v - \omega)] dt \right\} = \\ &= \frac{ev}{2\pi cr} e^{ikr} \left\{ \int_{-\infty}^0 \exp \left[i\omega t \left(\frac{v \epsilon^{\frac{1}{2}} \cos \theta}{c} - 1 \right) \right] dt + \int_0^{\infty} \exp \left[i\omega t \left(\frac{v \cos \theta}{c} - 1 \right) \right] dt \right\}. \end{aligned}$$

Making use of the definition of δ_+ and δ_- functions (see Appendix III, Vol. 1), we find

$$A_\omega = \frac{ev}{2\pi c\omega r} \frac{1}{i} \left\{ \frac{\delta_-}{1 - v\epsilon^{-1} \epsilon^{\frac{1}{2}} \cos \theta} - \frac{\delta_+}{1 - v\epsilon^{-1} \cos \theta} \right\}. \quad (39.25)$$

The intensity emitted according to formula (32.14) of Part I is equal to

$$\Delta E = \frac{e^2 \omega^2}{4\pi^2 c^2} \left[\frac{1}{1 - v c^{-1} \epsilon^{\frac{1}{2}} \cos \theta} - \frac{1}{1 - v c^{-1} \cos \theta} \right]^2 \sin^2 \theta \, d\Omega. \quad (39.26)$$

The meaning of this result is quite simple. In motion in a medium the role of the relative velocity v/c is played by the quantity $v\epsilon^{\frac{1}{2}}/c$, which determines the optical path in the medium. When the particle passes from one medium to another (or to vacuum) the quantity $v\epsilon^{\frac{1}{2}}/c$ changes discontinuously. This change is equivalent to an abrupt change in the velocity, i.e. to an acceleration of the particle. Formulae (39.25) and (39.26) show that the radiation is directed mainly forward ($\theta \approx 0$) and that it contains very high frequencies (high harmonics). We shall not dwell on the calculation of the total intensity, which can be accomplished by integrating over all angles. We note only that the total intensity turns out to be proportional to the energy of the particle. The analogous calculation for $v \ll c$ turns out to be more complicated, since in this case it is necessary to take into account the phenomena of reflection and refraction of electromagnetic waves at the interface.

Matter in the Plasma State

§40. The general characteristics of plasma

The problem of the passage of current through gases and the behaviour of gases conducting current in electric and magnetic fields is of great interest in present day physics.

As is well known, gases are normally non-conductors. If, however, a sufficiently large ionization is produced in the gas, i.e. if a sufficiently large number of free electrons and ions (positive and negative) are produced, the gas will become conducting. The process of the passage of current through the gas is called gaseous discharge. Depending on the mechanism causing the ionization of the gas the gaseous discharge is called non-self-maintained or self-maintained.

In the first case the basic ionization is produced by external sources (for example by γ -radiation or by a high temperature maintained by an external source).

In the self-maintained discharge the initial ionization is caused by electrons emerging from a cold cathode (the Townsend discharge and glow discharge).

The density of the current which can flow through the gas depends, in the first place, on the number of ions produced per cm^3 of the gas. In particular, if the entire gas is completely ionized current densities can be very large.

The self-maintained discharge displays a great diversity of properties. However, it possesses one outstanding feature: the space in which the gas discharge takes place can be divided into three regions — the cathode region, the anode region and the plasma region.

The properties of the cathode region and anode region depend on the mechanism of the discharge. In the cathode region the ionization of the atoms of the gas by electrons emerging from the cathode occurs. The greatest charge is concentrated in the cathode and anode regions and the major fall of the potential difference applied to the electrodes also takes place there. The size of these regions is, as a rule, not large, and they occupy only a small part of the space between the electrodes. The largest part of the interelectrode space is occupied by an ionized gas which is on the average electrically neutral. This region of the discharge is called a plasma. In a plasma the number of positive ions is on the average equal to the number of electrons and negative ions per unit volume. Besides ions and electrons, a plasma can also contain a larger or smaller number of neutral atoms or molecules.

The properties of a plasma in which we shall be interested in what follows do not depend on the detailed properties of the discharge nor on its character. The behaviour of a plasma plays an important part in the phenomena of gaseous discharge, a process which finds a wide variety of applications in modern engineering. A special interest in high-temperature plasmas has recently arisen in association with studies on controlled thermonuclear reactions as well as in connection with a number of astrophysical problems.

As is well known, in order to obtain thermonuclear reactions it is necessary to reach such high temperatures (higher than 10^8 degrees) that the energy of thermal motion of nuclear particles is sufficiently large to overcome the energy barriers which hinder the penetration of nuclei into each other. At such temperatures atoms are completely ionized and matter consists of a stripped plasma. Temperatures which are necessary for thermonuclear reactions to take place exist in the interiors of stars.

Under laboratory conditions it has not up to now been possible to prepare a plasma of the necessary temperature. However, intense investigations on high-temperature plasmas, which have already yielded a number of important results, are being carried out.

In astrophysical conditions matter is found in the plasma state not only in the interior of stars but also in stellar atmospheres and in clouds of interstellar matter.

§41. Equilibrium plasma

We naturally begin the study of the properties of a plasma with consideration of the theory of an equilibrium plasma.

We assume for simplicity that the plasma contains charges of only two kinds: positive ions with charge p_1 and electrons. In order to give a general character to the relations to be obtained we shall also call the latter ions and shall assign them the charge $p_2 = 1$. Then the condition of electrical neutrality for the plasma can be written in the form

$$\bar{n}_1 p_1 + \bar{n}_2 p_2 = 0, \quad (41.1)$$

where \bar{n}_1 and \bar{n}_2 are the mean numbers of ions and electrons respectively per unit volume.

In considering the equilibrium properties of a plasma we shall confine ourselves to the approximation of an ideal gas. In this approximation the Coulomb interaction between charged particles can be assumed to be small in comparison with the thermal energy:

$$p_1 p_2 e^2 / \bar{l} \ll kT, \quad (41.2)$$

where \bar{l} is the mean distance between the ions. The latter is connected with the number of ions per unit volume (the concentration of the plasma) N by the relation

$$\bar{l} \approx N^{-1/3}, \quad (41.3)$$

so that the condition of ideality of a gaseous plasma can be written in the form

$$N \ll \frac{kT}{p_1^3 p_2^3 e^6}. \quad (41.4)$$

The concentration of the plasma N is connected with the numbers \bar{n}_1 and \bar{n}_2 by the relation

$$N = \bar{n}_1 + \bar{n}_2. \quad (41.5)$$

If the inequality (41.4) is satisfied, then the plasma, in the zeroth order approximation, can be considered as an ordinary gas characterized by a temperature T .

The particles of the plasma will have a Maxwell velocity distribution and a uniform distribution in space. The Coulomb interaction between charged particles gives rise to a certain mean electric field in the volume of the plasma characterized by a potential $\bar{\varphi}$. In our approximation the change of the properties of the gas caused by this field can be assumed to be small.

In order to find the value of $\bar{\varphi}$ the following reasoning can be applied. We mentally single out a certain arbitrary ion in the plasma located at a point O which is chosen as the origin, and find the total mean potential of the electric field $\bar{\varphi}$ in the neighbourhood of the point O. The potential $\bar{\varphi}$ is produced by all ions (including the ion located at the point O). The averaging is carried out over all times of observation, during which the ions assume all possible positions in the plasma.

We consider a volume element dV located at a distance r from the origin O. Let the potential of the electric field in this volume be equal to $\bar{\varphi}(r)$. In view of the isotropy of the field, the potential $\bar{\varphi}$ depends only on the absolute value of r but not on the direction of the radius vector. For a low concentration of plasma one can apply the Boltzmann distribution law to the ions in the field, writing for the number of particles in the volume dV the expressions

$$n_1 dV = A \exp [-p_1 e \bar{\varphi} / kT] dV, \quad (41.6)$$

$$n_2 dV = B \exp [-p_2 e \bar{\varphi} / kT] dV, \quad (41.7)$$

where \bar{n}_1 and \bar{n}_2 are the numbers of positive and negative ions in unit volume.

The constants A and B can be found in the following way. At an arbitrarily high temperature $T \rightarrow \infty$ the field produced by the ions in the plasma cannot affect their spatial distribution, because their potential energy will be negligibly small. Hence as $T \rightarrow \infty$ both distributions must go over into a uniform distribution of the particles in space, i.e.

$$n_1 dV = \bar{n}_1 dV, \quad (41.8)$$

$$n_2 dV = \bar{n}_2 dV, \quad (41.9)$$

where n_1 and n_2 are the mean numbers of positive and negative ions per unit volume.

Comparing (41.8) and (41.9) with (41.6) and (41.7), we find

$$n_1 dV = \bar{n}_1 \exp [-p_1 e \bar{\varphi} / kT] dV, \quad (41.10)$$

$$n_2 dV = \bar{n}_2 \exp [-p_2 e \bar{\varphi} / kT] dV, \quad (41.11)$$

According to formulae (41.10) and (41.11), in the volume near the point O, where the ion which we have singled out is located, there is a charge

$$de = (\bar{n}_1 p_1 e \exp [-p_1 e \bar{\varphi} / kT] + \bar{n}_2 p_2 e \exp [-p_2 e \bar{\varphi} / kT]) dV. \quad (41.12)$$

This charge is determined by the fact that the probability of finding an ion in dV of the same sign as the ion at point O is somewhat lowered, whereas that of finding an ion of the opposite sign is somewhat increased, in comparison with the probability where we do not take into account the inter-ion interaction. In this sense it is said that a non-uniformly charged ion cloud arises around the ion located at point O.

It goes without saying that in practice there is no cloud around each of the ions, because we have singled out the ion at point O only for convenience of the argument, and all the ions in the plasma are equivalent. There is only a certain correlation probable between the positions of any pair of ions in space. This can be expressed in other words: it can be said that each ion produces around it an ion cloud and, at the same time, is a part of ion clouds of all the other ions in the plasma.

By means of (41.12) one can obtain the mean charge density at the point r :

$$\bar{\rho}(r) = de/dV = e(\bar{n}_1 p_1 \exp [-p_1 e \bar{\varphi} / kT] + \bar{n}_2 p_2 \exp [-p_2 e \bar{\varphi} / kT]). \quad (41.13)$$

Since by our assumption the energy of the inter-ion interaction is small in comparison with kT , one can expand the exponential expressions $\exp [-p_1 e \bar{\varphi} / kT]$ and $\exp [-p_2 e \bar{\varphi} / kT]$ in a series, writing

$$\bar{\rho}(r) \approx - \frac{e^2 (p_1^2 \bar{n}_1 + p_2^2 \bar{n}_2)}{kT} \bar{\varphi}(r). \quad (41.14)$$

The mean potential of the field $\bar{\varphi}$ at a given point of the plasma is connected with the mean charge density $\bar{\rho}$ at this point by the equation of electrostatics

$$\nabla^2 \bar{\varphi} = -4\pi \bar{\rho} / \epsilon. \quad (41.15)$$

Hence for $\bar{\varphi}$ we find the equation

$$\nabla^2 \bar{\varphi} = \kappa^2 \bar{\varphi}, \quad (41.16)$$

where κ^2 denotes the essentially positive quantity

$$\kappa^2 = \frac{4\pi e^2(p_1^2 \bar{n}_1 + p_2^2 \bar{n}_2)}{\epsilon k T}. \quad (41.17)$$

Eq. (41.16) or eq. (41.15), in which $\bar{\rho}$ is defined by (41.13), is called the Poisson-Boltzmann equation and represents the basis of the theory of an equilibrium plasma.

The solution of eq. (41.16) satisfying the requirement of isotropy in space can easily be obtained in polar coordinates. In polar coordinates, taking into account that $\bar{\varphi}$ does not depend on the polar angles θ and ψ , eq. (41.16) has the form

$$\frac{1}{r} \frac{d^2}{dr^2} (r\bar{\varphi}) = \kappa^2 \bar{\varphi}.$$

Introducing a new unknown function $f = r\bar{\varphi}$ we obtain

$$\frac{d^2 f}{dr^2} = \kappa^2 f.$$

The solution of the last equation has the form

$$f = C_1 e^{-\kappa r} + C_2 e^{\kappa r},$$

whence it follows that

$$\bar{\varphi} = C_1 r^{-1} e^{-\kappa r} + C_2 r^{-1} e^{\kappa r}. \quad (41.18)$$

The constant $C_2 = 0$, because the exponentially increasing solution leading to an infinitely large potential for $r \rightarrow \infty$ must be discarded. Hence

$$\bar{\varphi} = C_1 r^{-1} e^{-\kappa r}. \quad (41.19)$$

The constant C_1 can be found from the requirement that near the charge which is singled out the potential of the field should coincide with the Coulomb field of the charge. Hence it follows that

$$\varphi_{r \rightarrow 0} = C_1 r^{-1} \rightarrow p_1 e / \epsilon r,$$

so that

$$C_1 = p_1 e / \epsilon$$

and, finally,

$$\bar{\varphi} = \frac{p_1 e e^{-\kappa r}}{\epsilon r} . \quad (41.20)$$

Formula (41.20) shows that the potential of the field near the ion decreases, basically, according to an exponential law. At a distance $r > \kappa^{-1}$ from the ion the potential turns out to be small. The quantity κ^{-1} characterizing the rate of decrease of the potential is called the Debye length.

In order to elucidate the meaning of the solution obtained, we resolve the potential into the Coulomb potential of the ion which is singled out and the potential $\bar{\varphi}'$ of the field produced by all the other ions:

$$\bar{\varphi} = \frac{p_1 e}{\epsilon r} + \bar{\varphi}' .$$

From (41.20) we find

$$\bar{\varphi}' = \frac{p_1 e}{\epsilon} \frac{e^{-\kappa r} - 1}{r} . \quad (41.21)$$

Let us find the charge density corresponding to the potential $\bar{\varphi}'$. By virtue of (41.15) we have

$$\bar{\rho}' = -\frac{\epsilon}{4\pi} \nabla^2 \bar{\varphi}' = -\frac{p_1 e \kappa^2}{4\pi} \frac{e^{-\kappa r}}{r} .$$

This formula shows that near the ion with charge $p_1 e$ an ion cloud having the opposite sign of charge is formed. The charge density in the cloud decreases exponentially with the distance from the central ion. The total charge of the cloud is equal to

$$\int_0^\infty \bar{\rho}' dV = -p_1 e .$$

The meaning of this result is quite clear: around a given ion there are

grouped with a larger probability ions of the opposite sign. The total charge of the ion cloud surrounding any given ion is exactly equal to the charge of the given ion. The presence around an ion of a cloud of ions of the opposite sign leads to a weakening or, as is usually said, a screening of the field of the ion. Hence the potential of the screened field near the ion decreases more rapidly than the Coulomb potential. The quantity κ^{-1} represents the mean radius of the ion cloud surrounding a given ion.

Introducing the value of the Debye length $\kappa \sim e^2 N/kT$ into the condition of the applicability of the theory (41.4), we can rewrite it in the form

$$N\kappa^{-3} \gg 1,$$

i.e. in the form of the requirement that the mean number of ions confined in the volume of a sphere with a Debye radius must be large enough compared with unity.

The phenomenon of screening is of very great importance in the behaviour of a plasma. It is clear that any charge introduced into the plasma is screened at a distance of κ^{-1} .

Let, for example, the plasma be confined in a container with solid walls. If there is a surface charge on the walls then the field produced by it will be screened and will penetrate into the plasma only up to a depth κ^{-1} . The distance κ^{-1} is thus the thickness of the shielding layer which is formed at the boundary of an equilibrium plasma and which insulates it from external influences.

The same result can be obtained much more quickly by means of the method of correlation functions. That is, we make use of the smallness of the concentration of the plasma to close eq. (48.4) of Part III at the binary function. The latter contains the ternary function ρ_{123} . For small concentrations one can write approximately

$$\rho_{12}(r) = 1 + \psi_{12}(r), \quad (41.22)$$

where $\psi_{12}(r) \ll 1$. Formula (41.22) means that the interaction of particles in the plasma leads to a weak correlation. Further, if the probability of triple collisions between particles in the plasma is disregarded, then the ternary function ρ_{12j} can be written in the form of the product

$$\rho_{12j} = \rho_{12}\rho_{2j}\rho_{1j} \approx 1 + \psi_{12} + \psi_{2j} + \psi_{1j}. \quad (41.23)$$

Substituting this into (48.4) of Part III, we find,

$$\frac{\partial \psi_{12}}{\partial \mathbf{r}_1} = -\frac{1}{kT} \frac{\partial U_{12}(|\mathbf{r}_1 - \mathbf{r}_2|)}{\partial \mathbf{r}_1} - \frac{N}{VkT} \int \frac{\partial U_{1j}}{\partial \mathbf{r}_1} (1 + \psi_{12} + \psi_{2j} + \psi_{1j}) d\mathbf{r}_j. \quad (41.24)$$

We recall that the summation with respect to the index j is carried out over all (in our case two) kinds of particles.

We drop, as small, the term $\psi_{12}(\partial U_{12}/\partial \mathbf{r}_1)$. It is obvious that two integrals on the right-hand side reduce to zero:

$$\int \frac{\partial U_{12}(|\mathbf{r}_1 - \mathbf{r}_2|)}{\partial \mathbf{r}_1} d\mathbf{r}_j = 0,$$

$$\int \frac{\partial U_{1j}(|\mathbf{r}_1 - \mathbf{r}_j|)}{\partial \mathbf{r}_1} \psi_{2j}(|\mathbf{r}_2 - \mathbf{r}_j|) d\mathbf{r}_j = 0.$$

Indeed, they involve integration with respect to the angles of the vector $\partial U/\partial \mathbf{r}_1$, where U is an isotropic function of the corresponding variables. Hence, finally

$$\frac{\partial \psi_{12}}{\partial \mathbf{r}_1} = -\frac{1}{kT} \frac{\partial U_{12}(|\mathbf{r}_1 - \mathbf{r}_2|)}{\partial \mathbf{r}_1} - \frac{N}{VkT} \int \frac{\partial U_{1j}}{\partial \mathbf{r}_1} \psi_{2j} d\mathbf{r}_j. \quad (41.25)$$

We take the divergence of eq. (41.25) with respect to the coordinates \mathbf{r}_1 and take into account that the interaction is the Coulomb interaction, so that

$$\nabla \cdot \frac{\partial U_{12}}{\partial \mathbf{r}_1} = \nabla^2 U_{12}(|\mathbf{r}_1 - \mathbf{r}_2|) = 4p_1 p_2 \pi e^2 \delta(|\mathbf{r}_1 - \mathbf{r}_2|).$$

Then we find

$$\nabla^2 \psi_{12}(r) = \frac{4\pi p_1 p_2 e^2}{kT} \delta(r) + \frac{4\pi e^2}{VkT} p_1 N \sum p_j \psi_{1j}.$$

Setting

$$\psi_{12} = p_1 p_2 \psi(r),$$

we find finally

$$\nabla^2 \psi - \kappa^2 \psi = \frac{4\pi e^2}{kT} \delta(r), \quad (41.26)$$

which is the same as the equation for the mean potential (41.16). The term with the δ -function allows one to take the boundary condition (41.19) into account automatically.

It is easily seen that the correlation function satisfying (41.26) has the form

$$\psi_{12} = p_1 p_2 (1 - r^{-1} e^{-\kappa r}). \quad (41.27)$$

We shall now find the thermodynamic characteristics of an equilibrium plasma. The presence of the Coulomb interaction between the ions and electrons is responsible for the additional energy possessed by the plasma as compared to a neutral gas at the same pressure. This energy is obviously equal to $E' = \frac{1}{2} V \sum e n_i p_i \bar{\varphi}_i$, where n_i is the mean number of particles of the i th kind per unit volume, V is the total volume of the plasma, and $\bar{\varphi}_i$ is the mean potential produced by all ions at the locus of the i th ion.

The mean value of the potential of the electric field at distances $r < \kappa$ (for such distances the formulae derived above have a quantitative meaning) can be written in the form

$$\bar{\varphi}_i = -e p_i \kappa.$$

Hence

$$E' = -\frac{1}{2} e^2 V \kappa \sum n_i p_i^2 = -e^2 \left(\frac{\pi}{kTV} \right)^{\frac{1}{2}} \left(\sum n_i p_i^2 \right)^{\frac{3}{2}}.$$

Consequently, the total energy of the plasma is equal to

$$E = \sum n_i kTV - e^2 \left(\frac{\pi}{kTV} \right)^{\frac{1}{2}} \left(\sum n_i p_i^2 \right)^{\frac{3}{2}}.$$

Making use of the Gibbs-Helmholtz formula (30.11) of Part III, we find the free energy of the plasma

$$F = -T \int \frac{E}{T^2} dT = -V \sum n_i kT - \frac{2}{3} \left(\frac{\pi}{kTV} \right)^{\frac{1}{2}} \left(\sum n_i p_i^2 \right)^{\frac{3}{2}}.$$

The pressure of the plasma is equal to

$$p = - \left(\frac{\partial F}{\partial V} \right)_T = \sum n_i kT - \frac{e^2}{3V^{\frac{3}{2}}} \left(\frac{\pi}{kT} \right)^{\frac{1}{2}} \left(\sum n_i p_i^2 \right)^{\frac{3}{2}}.$$

The pressure of the plasma turns out to be lower than that of an ideal gas of the same density. This result has the following simple meaning: the attraction between charges of opposite signs, which are distributed closer to each other, predominates over the repulsion between charges of the same region.

In conclusion, we stress that although a plasma is a macroscopically homogeneous medium, on the scale $r < \kappa^{-1}$ it is inhomogeneous. This fact is of very great importance for electromagnetic processes in a plasma.

Up to now we have considered the plasma to be a completely equilibrium plasma. However, very often one has to study a plasma which is in incomplete equilibrium*. Namely, since the mass of the heavy ions is very large in comparison with the mass of the electrons, the energy exchange between them in elastic collisions proceeds very slowly. On the contrary, the energy exchange between electrons or between ions proceeds much more rapidly.

If at a certain initial instant of time the plasma was in a non-equilibrium state, then after the lapse of the relaxation time τ an equilibrium (Maxwell) distribution will be established for electrons and ions separately. Each set of particles can be characterized by the proper temperature T_e and T_i respectively. However, for the equalization of the temperatures and the establishment of a common temperature T of the plasma corresponding to equilibrium between the electrons and the ions, a relaxation time $\tau_T \gg \tau$ is necessary.

The presence of incomplete equilibrium in the plasma, which is characterized by two temperatures, does not have a very strong effect on the property of screening which we have described above.

§42. Plasma in static electric and magnetic fields

If a plasma is placed in an external static electric field E , then an electric current will arise in it.

In the absence of an external electric field the Maxwell velocity distribution is established for the ions and electrons in the plasma. When an external electric field is applied, the electrons and ions begin to move preferentially in different directions. A current arises in the plasma in the direction of the applied electric field. The density of this current is

* For more details on incomplete equilibrium see, for example, B.G. Levich, *Vvedenie v statisticheskuyu fiziku (Introduction to statistical physics)* (Gostekhizdat, Moscow, 1954).

$$\mathbf{j} = \sigma \mathbf{E}. \quad (42.1)$$

Up to now we have not tried to elucidate the meaning of the electric conductivity σ , assuming it to be a macroscopic characteristic of the medium. Here, however, it is necessary, if only on the basis of a very rough model, to estimate the value of σ . We proceed from the assumption that ions and electrons form an ideal gas. Let the mean velocities, masses and mean free paths of the ions and electrons be denoted by v_i , m_i , λ_i respectively. In the presence of an external electric field ions and electrons in the plasma undergo an acceleration between collisions in contrast to a neutral gas.

The mean velocity acquired by a particle under the action of the field \mathbf{E} is equal, in order of magnitude, to

$$\bar{u} \sim \frac{e_i}{m_i} E \tau_i,$$

where the mean time of flight between two successive collisions $\tau_i \sim \lambda_i/v_i$.

The systematic motion with velocity \bar{u} leads to a transport of charge in the direction of the field. The current density can be written in the form

$$\mathbf{j} = \sum n_i e_i \bar{u}_i \sim \left(\sum \frac{n_i e_i^2}{m_i} \tau_i \right) \mathbf{E}.$$

Thus, to within a numerical factor,

$$\sigma \sim \sum \frac{n_i e_i^2 \tau_i}{m_i}. \quad (42.2)$$

The values of τ_i will be found in Part VI of this book.

We assume that the value of τ is calculated or known from measurements. In weak fields and at relatively high pressures the electrical conductivity of a system consisting of electrons and randomly distributed ions has a constant value and does not depend on the applied field. The equilibrium velocity distribution of the electrons is disturbed only to a small degree by the external field.

For strong fields and low pressures the situation is different. Under the action of the applied field the electrons are accelerated (since the mean free path in a rarefied gas is sufficiently large) and acquire an energy which is considerably larger than the energy of thermal motion.

On the other hand, in collisions with ions the electrons lose energy. A calculation shows that a velocity distribution of the electrons can be estab-

lished, for which the increase in the energy of the electrons as they are accelerated in the field is compensated for by energy losses in collisions. Such a distribution is not an equilibrium distribution. Nevertheless, since it does not change in time (i.e. is stationary), one can speak of an effective temperature of the electrons which is equal to their mean energy. It turns out to be of the order of magnitude of

$$T_{\text{el}} \sim \frac{1}{k} \left(\frac{M_{\text{ion}}}{m_{\text{el}}} \right)^{\frac{1}{2}} \lambda_{\text{el}} eE, \quad (42.3)$$

(where M_{ion} and m_{el} are the masses of the ion and electron) and very large in comparison with the temperature of the ions and neutral molecules. The electrical conductivity is

$$\sigma \sim E^{-\frac{1}{2}}. \quad (42.4)$$

Let us now consider a plasma in constant electric and magnetic fields. As we shall see below, the magnetic field has a very important effect on the behaviour of the plasma.

The reasoning of §23 can be applied to a system of electrons moving with a velocity \mathbf{v} . Namely, when electrons are moving with respect to the magnetic field an electric field of strength

$$\mathbf{E}_{\text{ind}} = \frac{\mu}{c} (\mathbf{v} \times \mathbf{H}) \quad (42.5)$$

is induced, where μ is the magnetic permeability of the medium (the plasma). In the presence of an electric field \mathbf{E} a current

$$\mathbf{j} = \sigma(\mathbf{E} + \mathbf{E}_{\text{ind}}) = \sigma \left(\mathbf{E} + \frac{\mu}{c} (\mathbf{v} \times \mathbf{H}) \right) = \sigma \mathbf{E} + \frac{\sigma \mu}{c} (\mathbf{v} \times \mathbf{H}) \quad (42.6)$$

arises in the system of electrons. Expressing σ in terms of the quantities involved in it and noting that $\mathbf{j} = n_{\text{el}} e \mathbf{v}$, we can rewrite this last formula in another form:

$$\mathbf{j} = \sigma \mathbf{E} + \alpha (\mathbf{j} \times \mathbf{H}), \quad (42.7)$$

where

$$\alpha = \frac{e\mu\tau_{el}}{m_{el}c} = \frac{\omega_H\mu\tau_{el}}{H} \quad (42.8)$$

and ω_H is the cyclotron frequency for the electrons.

It should be noted that formula (42.7) also remains valid when one takes into account the velocity distribution of the electrons since this has an effect only on the numerical coefficient α .

Let us consider two cases:

- (1) the electric field parallel to the magnetic field, $\mathbf{E} \parallel \mathbf{H}$,
- (2) the electric field perpendicular to the magnetic field, $\mathbf{E} \perp \mathbf{H}$.

In the first case the magnetic field does not directly affect the current density $\mathbf{j} = \sigma\mathbf{E}$ so that $\mathbf{j} \times \mathbf{H} = 0$.

The second case is of great interest. To find the vector \mathbf{j} from (42.7) we first form the scalar product of (42.7) and \mathbf{H} and then the vector product of (42.7) and \mathbf{H} . We then have

$$\mathbf{j} \cdot \mathbf{H} = \sigma(\mathbf{E} \cdot \mathbf{H}) + \alpha[(\mathbf{j} \times \mathbf{H}) \cdot \mathbf{H}] = + \alpha[(\mathbf{j} \times \mathbf{H}) \cdot \mathbf{H}] ,$$

or, since $(\mathbf{j} \times \mathbf{H}) \cdot \mathbf{H} = \mathbf{j} \cdot (\mathbf{H} \times \mathbf{H}) = 0$, we obtain

$$\mathbf{j} \cdot \mathbf{H} = 0 . \quad (42.9)$$

Further,

$$\begin{aligned} \mathbf{j} \times \mathbf{H} &= \sigma(\mathbf{E} \times \mathbf{H}) + \alpha[(\mathbf{j} \times \mathbf{H}) \times \mathbf{H}] = \\ &= \sigma(\mathbf{E} \times \mathbf{H}) + \alpha\mathbf{H}(\mathbf{j} \cdot \mathbf{H}) - \alpha\mathbf{j}H^2 = \sigma(\mathbf{E} \times \mathbf{H}) - \alpha\mathbf{j}H^2 . \end{aligned} \quad (42.10)$$

Whence, substituting (42.10) into (42.7), we find

$$\mathbf{j} = \sigma\mathbf{E} + \alpha\sigma(\mathbf{E} \times \mathbf{H}) - \alpha^2H^2\mathbf{j}$$

or

$$\mathbf{j} = \frac{\sigma}{1 + \alpha^2H^2} \mathbf{E} + \frac{\sigma\alpha}{1 + \alpha^2H^2} (\mathbf{E} \times \mathbf{H}) . \quad (42.11)$$

Formula (42.11) shows that in the presence of the magnetic field the current is no longer parallel to the electric field. Ohm's law in its ordinary form does not hold.

From formula (42.11) it follows that in the direction of the electric field the current density is

$$j_{\parallel} = \frac{\sigma}{1 + \alpha^2 H^2} E = \sigma_{\parallel} E. \quad (42.12)$$

The electrical conductivity σ_{\parallel} of the plasma in the case of $\mathbf{H} \perp \mathbf{E}$ turns out to be smaller by a factor of $1 + \alpha^2 H^2$ than the electrical conductivity in the absence of the magnetic field.

Besides the current in the direction of the electric field there arises a current j_{\perp} in the direction perpendicular to \mathbf{E} and \mathbf{H} ; this current is equal in absolute value to

$$j_{\perp} = \frac{\sigma \alpha H}{1 + \alpha^2 H^2} E = \sigma_{\perp} E. \quad (42.13)$$

The current j_{\perp} is called the Hall current, and $\sigma_{\perp} = \sigma_{\parallel} \alpha H$ is called the Hall conductivity. For $\alpha H = \mu \tau \omega_H \gg 1$ the Hall current can substantially exceed the ordinary current. The condition $\tau_{cl} \omega_H \gg 1$ can be written in the obvious form

$$\tau_{cl} \omega_H = \frac{eH}{mc} \frac{\lambda}{v} \sim \frac{\lambda}{R},$$

where $R = mc v / eH$ is the radius of the circle described by an electron in the magnetic field \mathbf{H} . The condition $\tau_{cl} \omega_H \gg 1$ is fulfilled in a strong magnetic field and a rarefied plasma.

The effect of the magnetic field on the current in the plasma and, in particular, the appearance of the current component j_{\perp} has a simple meaning. It is associated with the character of motion of a charge in crossed electric and magnetic fields as described in §39 of Part I. In such fields particles undergo a drift whose velocity and direction are determined by formula (39.15) of Part I.

The motion of charges in the direction perpendicular to \mathbf{E} and \mathbf{H} leads to the appearance of a current with the density j_{\perp} . When the reverse inequality $\omega \tau_{cl} \ll 1$ is fulfilled the effect of the magnetic field on the conductivity becomes small.

We shall not dwell on the differences in the above expressions associated with the motion of positive ions. Since the mass of the ions is large, the inequality $\omega \tau_{ion} \ll 1$ is usually fulfilled for them and the magnetic field has no significant effect on the ion part of the electrical conductivity.

§43. Magnetic isolation and the pinch effect

Electric and magnetic forces acting on a plasma give rise to important mechanical effects. Conversely, a spatial displacement of the plasma has a significant effect on its behaviour in electromagnetic fields.

In order to consider the mutual effect of the field and the mechanical motion of the plasma we shall write the general equations which describe the properties of the plasma.

The electromagnetic field in the plasma is described by Maxwell's equations. The equations of motion of the plasma are the equations of hydrodynamics*. Disregarding the viscosity of the plasma, one can write the equations of motion in the form

$$\delta_0 \frac{d\mathbf{v}}{dt} = \mathbf{F} - \nabla p,$$

where \mathbf{F} is the ponderomotive force acting on unit volume of the plasma, δ_0 is the density and p is the pressure of the gas.

If a current \mathbf{j} is flowing in the plasma, then in the magnetic field

$$\mathbf{F} = \frac{\mu}{c} (\mathbf{j} \times \mathbf{H}). \quad (43.1)$$

Hence

$$\delta_0 \frac{d\mathbf{v}}{dt} = -\nabla p + \frac{\mu}{c} (\mathbf{j} \times \mathbf{H}). \quad (43.2)$$

In particular, for a motionless plasma ($\mathbf{v}=0$) one can write the equation of hydrostatics:

$$\nabla p = \frac{\mu}{c} (\mathbf{j} \times \mathbf{H}). \quad (43.3)$$

We shall apply the equation of hydrostatics to the consideration of the important phenomenon of the magnetic isolation of a plasma.

* See, for example, L.G.Loytsyanskii, *Mekhanika zhidkosti i gazov (Mechanics of liquids and gases)* (Gostekhizdat, Moscow, 1950); L.D.Landau and E.M.Lifshitz, *Mechanics of continuous media* (Pergamon Press, Oxford, 1960).

Let us consider the plasma in a magnetic field \mathbf{H} perpendicular to the current vector \mathbf{j} . Then from eq. (43.3) it is clear that $\nabla p \neq 0$ and the pressure of the plasma varies from point to point. Eq. (43.3) can be integrated if the current density is eliminated from it by means of Maxwell's equations. We have from (43.3) and (14.3)

$$\nabla p = \frac{\mu}{4\pi} [(\nabla \times \mathbf{H}) \times \mathbf{H}] . \quad (43.4)$$

By means of the vector equality (1.48) we have

$$(\nabla \times \mathbf{H}) \times \mathbf{H} = (\mathbf{H} \cdot \nabla) \mathbf{H} - \nabla \left(\frac{1}{2} H^2 \right) .$$

Hence

$$\nabla p = - \frac{\mu}{4\pi} \nabla \left(\frac{1}{2} H^2 \right) + \frac{\mu}{4\pi} (\mathbf{H} \cdot \nabla) \mathbf{H} . \quad (43.5)$$

Let us choose the direction of the field as the x -axis and consider the particular case where the magnetic field strength does not vary in the longitudinal direction (i.e. $\partial \mathbf{H} / \partial x = 0$) but can vary according to an arbitrary law as a function of the coordinates y, z . In other words, we consider a field with strength $\mathbf{H} = \mathbf{H}(y, z)$ which varies in space. In this case we have, obviously,

$$(\mathbf{H} \cdot \nabla) \mathbf{H} = H \frac{\partial \mathbf{H}(y, z)}{\partial x} = 0$$

and formula (43.5) gives

$$p + \frac{\mu H^2}{8\pi} = \text{const} . \quad (43.6)$$

The quantity $\mu H^2 / 8\pi$ represents the magnetic pressure, i.e. the force acting on a unit area of an imaginary plane in the gas.

Formula (43.6) shows that the total pressure of the plasma, made up of the magnetic pressure $\mu H^2 / 8\pi$ and the gas pressure p , must remain constant in space.

As an example, let the plasma which is in a magnetic field not be bounded by impenetrable walls. Then the relation (43.6) shows that the total pressure

cannot drop to zero at any point in the plasma. In the region of space which is not occupied by the plasma the value of \mathbf{H} is larger than in the inner region occupied by the plasma. This means that the plasma cannot expand into a vacuum. The magnetic field isolates the plasma, replacing the impenetrable wall.

Another important hydrostatic effect is the pinch effect or the phenomenon of a plasma column. This phenomenon consists in a compression of the plasma by the magnetic field of the plasma current itself.

Let the plasma be represented by a cylinder of radius R (we direct the axis of the cylinder along the z -direction) along which a current of density j is flowing. The magnetic field of the current produces a magnetic pressure which must be balanced by the pressure of the plasma. The pressure of the plasma is most simply found by assuming the current density j to be constant over its cross-section (i.e. assuming that $j = j_0$ for $r < R$, and $j = 0$ for $r > R$). An analogous result can be obtained in the general case by means of formula (43.5). Then in cylindrical coordinates (43.3) assumes the form

$$\frac{dp}{d\rho} = -\frac{jH_\psi}{c},$$

where the magnetic field H_ψ is expressed by formula (17.11).

Integration gives

$$p = p_0 - \pi j^2 \rho^2 / c^2 \quad (\rho < R), \quad (43.7)$$

$$p = 0 \quad (\rho > R). \quad (43.8)$$

Here $p_0 = nkT$ is the pressure and n_0 is the density of the gas at the centre of the plasma cylinder.

Formulae (43.7) and (43.8) show that the gas pressure and correspondingly the density of the gas at the centre are higher than at the periphery of the cylinder. The magnetic field of the plasma current compresses and maintains the plasma cylinder. The radius of the plasma cylinder has a constant value, and the release of Joule heat leads to its heating.

The phenomenon of the self-constriction of a plasma cylinder, which is called the "pinch effect", leads to the detachment of the plasma from the walls of the container in which the gaseous discharge takes place and to the formation of a more or less slender plasma column. A simple spark or lightning discharge are examples of a plasma column. The formation and constriction of a plasma column is, naturally, of particularly great importance for large current densities.

We have confined ourselves here to finding the pressure distribution only on the assumption that the behaviour of the plasma has a stationary character. In practice, the non-stationary motion of the plasma, which leads to oscillations of the plasma cylinder as a whole, and which can lead to the loss of stability and rupture, is of importance in bringing about the pinch effect.

The study of the whole picture of non-stationary phenomena arising in the pinch effect is very complex *.

§44. The magnetic field in a moving plasma

In a number of important problems hydrodynamical effects associated with the macroscopic motion of the plasma play an essential role in the behaviour of the plasma. For the study of such effects it is necessary to formulate a system of equations for the electromagnetic field in a moving medium.

On the basis of the results of §23 one can write

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} + \nabla \times (c^{-1} \mathbf{v}_0 \times \mathbf{B}). \quad (44.1)$$

Moreover, if the fields are varying sufficiently slowly in time and if the displacement current can be disregarded, the magnetic field distribution is defined by eqs. (22.5). By means of Ohm's law we write (44.1) in the form

$$\frac{\partial \mathbf{B}}{\partial t} = -c(\nabla \times \mathbf{E}) + \nabla \times (\mathbf{v}_0 \times \mathbf{B}) = -c(\nabla \times \sigma^{-1} \mathbf{j}) + \nabla \times (\mathbf{v}_0 \times \mathbf{B}). \quad (44.2)$$

Eliminating the current density from (44.2) and (22.5), we have

$$\frac{\partial \mathbf{B}}{\partial t} = -\frac{c^2}{4\pi\sigma\mu} \nabla \times (\nabla \times \mathbf{B}) + \nabla \times (\mathbf{v}_0 \times \mathbf{B}).$$

* For problems concerning the plasma state of matter see L. Spitzer, *Physics of fully ionized gases* (Interscience Publ., New York, 1956 and 1962); T. Cowling, *Magnetohydrodynamics* (Interscience Publ., New York, 1957); H. Alfvén, *Cosmical electrodynamics* (Clarendon Press, Oxford, 1953); collection of papers, *Upravlyemyye termoyadernyye reaktsii* (Controlled thermonuclear reactions), (Atomizdat, Moscow, 1960); L. A. Artsimovich, *Upravlyayemye termoyadernyye reaktsii* (Controlled thermonuclear reactions), (Fizmatgiz, Moscow, 1961).

According to (1.50) and taking into account (22.5), we find finally

$$\frac{\partial \mathbf{B}}{\partial t} = \frac{c^2}{4\pi\sigma\mu} \nabla^2 \mathbf{B} + \nabla \times (\mathbf{v}_0 \times \mathbf{B}). \quad (44.3)$$

Comparing (44.3) with (30.1) we see that, in contrast to the case of a motionless medium, eq. (44.3) contains the term $\nabla \times (\mathbf{v}_0 \times \mathbf{B})$. In the absence of this term eq. (44.3) expresses the attenuation of the magnetic field in the conducting medium over the depth of the skin layer.

We shall be more interested here in the case where the first term on the right-hand side of (44.3) can be neglected in comparison with the second term. For this it is necessary that the velocity of motion \mathbf{v}_0 and the conductivity σ of the plasma be sufficiently large*.

This situation is relatively difficult (although possible) to attain in laboratory conditions. However, it is just this case which is dealt with in studying phenomena occurring on a cosmic scale. Dropping the small term in eq. (44.3), one can rewrite it in the form

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v}_0 \times \mathbf{B}). \quad (44.4)$$

Relation (44.4) has an important meaning. According to (22.6), eq. (44.4) means that the flux of magnetic induction through a closed contour, every point of which is moving along with the fluid, is constant in time. One can picture this condition in an obvious way by means of the lines of force of the magnetic field. The equality (44.4) means that the field lines are moving together with the bulk of the plasma, as if frozen in to it.

Let us consider a closed "fluid contour", i.e. a closed contour connecting particles of the fluid, each of which is moving along its own hydrodynamical flow line. From (44.4) it follows that the number of lines of the magnetic field which cross the fluid contour remains constant. The fluid particles move as if sliding along the field lines and do not intersect them in the transverse direction.

Let us now consider the motion of the plasma perpendicular to the magnetic field. One can picture the peculiarities of such a motion most simply by considering a simple case. At the initial instant of time let the plasma be at rest and then set in motion with the velocity profile shown by arrows in

* A more precise formulation can be found in the monographs of Spitzer or Cowling which we have already mentioned.

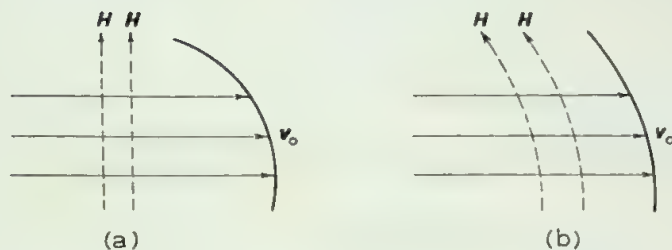


Fig. IV.19

fig. IV.19. The lines of magnetic field strength in the plasma at rest are shown in fig. IV.19a by dashed lines. The motion of the plasma "carrying along" the lines gives them the form which is shown in fig. IV.19b by dashed lines.

In the plasma at rest the magnetic field has the strength $H_x^{(0)}$. In the moving plasma besides the field component H_x the component $H_y \neq 0$ arises. It is easily shown that as the lines of the magnetic field are deformed its strength increases.

Let the equation of a field line deformed by the motion be $y(x)$. We assume that the bending of the field line is small. Then we can write

$$\frac{dy}{dx} = \frac{H_y}{H_x} \approx \frac{H_y}{H_x^{(0)}},$$

whence

$$H_y = H_x^{(0)} \frac{dy}{dx}.$$

Let us find the change in the energy of the magnetic field under such a deformation of the field lines. In order to make the formulae to be obtained clearer we shall relate this energy to one line of the magnetic field. For this we note that, by definition, H lines of the magnetic field pass through unit area perpendicular to the field. If

$$U = \frac{\mu}{8\pi} \int H^2 dx dS = \frac{\mu}{8\pi} \int (H_x^{(0)})^2 dx dS = \mu \frac{(H_x^{(0)})^2}{8\pi} S \int dx$$

is the total initial energy of the field, then the energy per unit area is $U/S = (\mu/8\pi)(H_x^{(0)})^2 \int dx$ and, correspondingly, the energy per line is equal to

$$w_0 = \frac{U}{SH_x^{(0)}} = \frac{1}{8\pi} \mu H_x^{(0)} \int dx.$$

After the deformation this quantity can be written in the form

$$\begin{aligned} w &= \frac{\mu}{8\pi H_x^{(0)}} \left\{ \int [H_x^2 + H_y^2] dx \right\} \approx \\ &\approx \frac{\mu}{8\pi H_x^{(0)}} \left\{ \int (H_x^{(0)})^2 dx + \int (H_x^{(0)})^2 \left(\frac{dy}{dx} \right)^2 dx \right\} = \\ &= w_0 + \frac{\mu H_x^{(0)}}{8\pi} \int \left(\frac{dy}{dx} \right)^2 dx. \end{aligned}$$

The increase in the energy of the magnetic field per line is equal to

$$\Delta w = w - w_0 = \frac{\mu H_x^{(0)}}{8\pi} \int \left(\frac{dy}{dx} \right)^2 dx. \quad (44.5)$$

This increase in the energy is due to the work done by the moving fluid against the elastic force of resistance of the field line.

It is interesting to compare the expression obtained with the potential energy of a deformed elastic string. If the tension of the string is denoted by α , then the latter quantity can be written in the form

$$\Delta U = \alpha \int \{ [1 + (dy/dx)^2]^{\frac{1}{2}} - 1 \} dx,$$

where $[1 + (dy/dx)^2]^{\frac{1}{2}} - 1$ is the geometrical elongation of the string under deformation. Assuming the deflection to be small, we have *

$$\Delta U \approx \frac{1}{2} \alpha \int (dy/dx)^2 dx. \quad (44.6)$$

The comparison of (44.6) with (44.5) shows that the line of magnetic field strength in the plasma behaves as a string with effective tension

* See, for example, A.N.Tikhonov and A.A.Samarskii, *Partial differential equations of mathematical physics* (Holden Day, San Francisco, 1964).

$$\alpha_{\text{eff}} = \mu H_x^{(0)} / 4\pi. \quad (44.7)$$

If the deformation of the lines of the field cannot be considered as small, the expression derived above for the increase in the energy of the field turns out to be inadequate. However, the general result remains valid: the deformation and elongation of the lines of magnetic field strength which are carried along by the moving fluid correspond to an increase in the strength of the field.

Thus, the motion of the conducting liquid can, in principle, lead to the generation and intensification of the magnetic field. On the other hand, if the liquid is placed in a sufficiently strong magnetic field, then this field can hinder the motion of the fluid, which is as if it attached to the magnetic field. The magnetic field also hampers the transition from laminar motion of a conducting fluid to turbulent motion.

It turns out that these results can be checked directly in laboratory experiments.

Certain important consequences of the properties of the "frozen-in" magnetic field described above will be considered in the next section.

§45. Magnetohydrodynamic waves

The analogy between the properties of the elastic string and the lines of the magnetic field naturally suggests the idea of the possibility of appearance of oscillations of the magnetic field about a certain equilibrium configuration.

Let us consider a plasma fluid in a magnetic field of strength \mathbf{H}_0 . We assume that the field \mathbf{H}_0 is uniform and constant in time. Let an infinitesimally small perturbation in the form of a field with velocity \mathbf{v} arise in the fluid. We assume that the conductivity of the plasma is infinitely large, so that the motion of the plasma fluid completely carries along the lines of magnetic field strength.

Then in the velocity field the strength of the magnetic field can be written in the form

$$\mathbf{H} = \mathbf{H}_0 + \mathbf{h}, \quad (45.1)$$

where $\mathbf{h} = \mathbf{h}(\mathbf{r}, t)$ is an infinitesimally small perturbation. Substituting (45.1) into eq. (44.4) and disregarding the product of the infinitesimally small quantities, we have

$$\frac{\partial \mathbf{h}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{H}_0) = (\mathbf{H}_0 \cdot \nabla) \mathbf{v}. \quad (45.2)$$

Here we have made use of formula (I.45) and the condition of incompressibility of the fluid:

$$\nabla \cdot \mathbf{v} = 0. \quad (45.3)$$

The equation of motion of the fluid (43.2) can be simplified, noting that in our case of an infinitesimally small velocity one can write

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \approx \frac{\partial \mathbf{v}}{\partial t}$$

and that the ponderomotive force can be written, as in deriving (43.5), in the form

$$\mathbf{F} = \frac{\mu}{c} (\mathbf{j} \times \mathbf{H}) = \frac{\mu}{4\pi} [(\nabla \times \mathbf{H}) \times \mathbf{H}] \approx \frac{\mu}{4\pi} [(\nabla \times \mathbf{h}) \times \mathbf{H}_0],$$

with an accuracy to within the second order of small quantities. According to formula (I.47) we have

$$(\nabla \times \mathbf{h}) \times \mathbf{H}_0 = (\mathbf{H}_0 \cdot \nabla) \mathbf{h} - \nabla (\mathbf{H}_0 \cdot \mathbf{h}),$$

taking into account that \mathbf{H}_0 is a constant vector. Hence the equation of motion of the field (43.2) reads:

$$\delta_0 \frac{\partial \mathbf{v}}{\partial t} = -\nabla \left(p + \frac{\mu (\mathbf{H}_0 \cdot \mathbf{h})}{4\pi} \right) + \frac{\mu}{4\pi} (\mathbf{H}_0 \cdot \nabla) \mathbf{h}. \quad (45.4)$$

Without restricting the general character of our reasoning, we can choose the direction of the unperturbed field \mathbf{H}_0 as the x -axis. Then eqs. (45.2) and (45.4) can be rewritten in the form

$$\frac{\partial \mathbf{h}}{\partial t} = H_0 \frac{\partial \mathbf{v}}{\partial x}, \quad (45.5)$$

$$\delta_0 \frac{\partial \mathbf{v}}{\partial t} = -\nabla \left(p + \frac{\mu (\mathbf{H}_0 \cdot \mathbf{h})}{4\pi} \right) + \frac{\mu}{4\pi} H_0 \frac{\partial \mathbf{h}}{\partial x}. \quad (45.6)$$

It is easy to show that the solution of the system of eqs. (45.5) and (45.6) represents a system of plane waves propagating along the x -axis (in the direction of the unperturbed field).

From conditions (45.3) and (45.4) it follows that such waves must be transversal. Choosing the vector \mathbf{h} as the y -axis and taking components of eqs. (45.5) and (45.6) in the coordinate axes, we have

$$h_x = 0, \quad (45.7)$$

$$\frac{\partial h_y}{\partial t} = H_0 \frac{\partial v_y}{\partial x}, \quad (45.8)$$

$$h_z = 0, \quad (45.9)$$

$$v_x = 0, \quad (45.10)$$

$$\delta_0 \frac{\partial v_y}{\partial t} = \frac{\mu}{4\pi} H_0 \frac{\partial h_y}{\partial x}, \quad (45.11)$$

$$v_z = 0. \quad (45.12)$$

Here we have assumed that the pressure also depends only on the coordinate x and that the vector ∇p has neither a y -component nor a z -component.

Differentiating (45.8) with respect to t and taking into account (45.11), we obtain

$$\frac{\partial^2 h_y}{\partial x^2} - \frac{1}{c_m^2} \frac{\partial^2 h_y}{\partial t^2} = 0. \quad (45.13)$$

Analogously, differentiating (45.11) with respect to t and taking into account (45.8), we have

$$\frac{\partial^2 v_y}{\partial x^2} - \frac{1}{c_m^2} \frac{\partial^2 v_y}{\partial t^2} = 0, \quad (45.14)$$

where

$$c_m^2 = \frac{\mu H_0^2}{4\pi\delta_0}. \quad (45.15)$$

Eqs. (45.13) and (45.14) show that the perturbation which arose in the plasma propagates in the form of plane waves along the magnetic field \mathbf{H}_0 with a constant velocity c_m determined by formula (45.15):

$$v_y = a e^{i(kx - \omega t)}, \quad (45.16)$$

$$h_y = b e^{i(kx - \omega t)}. \quad (45.17)$$

These waves are called the Alfvén magnetohydrodynamic waves.

We assume for brevity of notation that these waves are monochromatic and that they are propagated in the positive direction of the x -axis. An outstanding property of magnetohydrodynamic waves is the fact that they propagate in an incompressible fluid.

As is well known, only sound waves, associated with a change in the density of the medium, can propagate in a non-conducting fluid.

The transverse character of magnetohydrodynamic waves is clear from the foregoing. The velocity of their propagation is determined by the properties of the medium and the strength of the constant magnetic field \mathbf{H}_0 . The relation between the amplitude of the magnetic field and the velocity is obtained by substituting (45.16) and (45.17) into the initial equations:

$$|v_y| = \left(\frac{\mu}{4\pi\delta_0} \right)^{\frac{1}{2}} |h_y|.$$

It goes without saying that besides the magnetic field there is also an electric field in the magnetohydrodynamic wave, defined by the Maxwell equation (4.13):

$$E_z = |h_y| \frac{H_0 \mu^{\frac{3}{2}}}{c(4\pi\delta_0)^{\frac{1}{2}}} e^{i(kx - \omega t)}.$$

It is useful to note that magnetohydrodynamic waves can be obtained from obvious considerations associated with the analogy between the lines of magnetic field and an elastic string. As is well known*, the equation of motion of the string has the form

* See, for example, A.N.Tikhonov and A.A.Samarskii, *Partial differential equations of mathematical physics* (Holden Day, San Francisco, 1964).

$$\frac{\partial^2 \xi}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 \xi}{\partial t^2} = 0.$$

where ξ is the transverse displacement, and c is the velocity of propagation of waves along the string equal to $c = (\alpha/\delta_0)^{1/2}$. If one substitutes for α its effective value (44.7), then c is the same as the velocity of propagation of magneto-hydrodynamic waves.

Magnetohydrodynamic waves are essentially a special case of electromagnetic waves in a conducting medium. In the presence of a sufficiently strong magnetic field in the conducting fluid a strong damping of electromagnetic waves occurs for propagation in all directions except in the direction of the magnetic field. Thus, the conducting fluid then has a strongly pronounced anisotropy for electromagnetic properties.

If the velocity of propagation of magnetohydrodynamic waves is written in the form $c_m = c_{\text{eff}}/(\epsilon\mu)^{1/2}$ where c_{eff} is the velocity replacing the velocity of light in the analogous formula, then the dielectric constant of the conducting fluid turns out to be

$$\epsilon = \frac{4\pi c_{\text{eff}}^2 \delta_0}{\mu^2 H_0^2}.$$

For $c_{\text{eff}} \approx 100$ cm/sec, which corresponds to the values of the field $H_0 \approx 300$ gauss and to $\delta_0 \approx 1$, ϵ is of the order of magnitude of 3×10^{18} .

Magnetohydrodynamic waves, as do all electromagnetic waves, transport energy. The energy flux transported propagates with a velocity c_m which, for a sufficiently strong field H_0 , is very large in comparison with the velocity of motion $|v|$ of the matter in a plasma fluid.

Thus, in contrast to an ordinary non-conducting fluid, for a plasma in a magnetic field H_0 there is always a mechanism ensuring a rapid transport of the energy of a perturbation arising in the plasma.

The properties of the plasma state described above determine the behaviour of matter at high temperatures, when atoms are to a considerable degree ionized. Hence the study of the plasma state is important, on the one hand, for astrophysics and, on the other hand, for investigations in the field of controlled thermonuclear reactions.

We cannot in this book discuss the vast fields of investigation mentioned above, and for this we refer the reader to the specialized literature.

§46. Plasma in a high-frequency electric field

In considering the behaviour of plasma in stationary and quasistationary fields, we have disregarded displacement currents and have assumed a plasma to be a homogeneous conducting fluid. This approximation of magneto-hydrodynamics turns out, however, to be unsatisfactory in the realm of high-frequency processes.

The substantial difference between the masses of electrons and heavy ions has an important effect in high-frequency fields. In this case the so-called two-fluid model often proves to be a sufficiently good approximation.

In the two-fluid model approximation (it would be more appropriate to call it the gas-mixture model) the ions and electrons are assumed to be two ideal gases moving independently of each other under the action of corresponding forces.

For electrons and ions separately one can write equations of motion — the equations of hydrodynamics:

$$mn \frac{d\mathbf{v}}{dt} = -\nabla p + en\mathbf{E}. \quad (46.1)$$

All the quantities, the charge, the mass, and the number of particles n per cm^3 , are referred to the electrons or the ions. For brevity in the notation we drop the index characterizing the type of particle. The continuity equation is also written separately for electrons and ions

$$\frac{\partial n}{\partial t} + \nabla \cdot n\mathbf{v} = 0. \quad (46.2)$$

We assume the pressure in the plasma for each kind of particle to be equal to the gas pressure

$$p = nkT. \quad (46.3)$$

Finally, we assume that the oscillations performed by the charges under the action of the field \mathbf{E} are undamped.

The problem of damping will be considered in detail in Part VI, where it will be shown that in practice dissipative processes, which play an important part in a number of phenomena, occur in plasma. However, here we can disregard dissipative processes and assume the oscillations to be adiabatic, so that the pressure and density are interrelated by the equation of an adiabatic curve.

Taking this into account, one can write for ∇p in (46.1)

$$p = \left(\frac{\partial p}{\partial n} \right)_S \nabla n,$$

so that

$$m n \frac{d\mathbf{v}}{dt} = - \left(\frac{\partial p}{\partial n} \right)_S \nabla n + e n \mathbf{E}.$$

We assume that the field \mathbf{E} , the field of the electromagnetic wave in the plasma, varies according to the law $\mathbf{E} \sim \exp [i(\mathbf{k} \cdot \mathbf{r} - \omega t)]$. Also, we assume that the charges perform small oscillations under the action of the field. The changes in the density of the plasma which occur in this case are small, and one can write that

$$n = n_0 + n', \quad n' \ll n_0,$$

where n_0 is the density of the plasma in the absence of an external field.

We look for a solution of the above system of equations in the form

$$v \sim n' \sim p' \sim \exp [i(\mathbf{k} \cdot \mathbf{r} - \omega t)], \quad (46.5)$$

disregarding the squares of these quantities. We then have

$$\frac{\partial n'}{\partial t} + n_0 (\nabla \cdot \mathbf{v}) = 0, \quad (46.6)$$

$$m n_0 \frac{d\mathbf{v}}{dt} = - \left(\frac{\partial p}{\partial n} \right)_S \nabla n' + e n_0 \mathbf{E}. \quad (46.7)$$

Substituting the exponential functions into (46.6) and (46.7), we find

$$n' = \frac{n_0}{\omega} (\mathbf{k} \cdot \mathbf{v}), \quad (46.8)$$

$$\mathbf{v} = i \frac{e \mathbf{E}}{m \omega} + \left(\frac{\partial p}{\partial n} \right)_S \frac{\mathbf{k} (\mathbf{k} \cdot \mathbf{v})}{m \omega^2}. \quad (46.9)$$

If the direction of the vector \mathbf{k} is chosen as the z -axis, then for an iso-

tropic plasma the velocity vector can be written in the form $\mathbf{v} = (v_{\parallel}, \mathbf{v}_{\perp})$, where v_{\parallel} is the component of the vector in the direction of propagation of the wave, and \mathbf{v}_{\perp} is the velocity vector in the xy -plane. From (46.9) we have

$$\begin{aligned} \mathbf{v}_{\perp} &= \frac{ie}{m\omega} \mathbf{E}_{\perp}, \\ v_{\parallel} &= \frac{ieE_z}{m\omega} + \left(\frac{\partial p}{\partial n}\right)_S \frac{k^2}{m\omega^2} v_{\parallel}, \end{aligned} \quad (46.10)$$

or

$$v_{\parallel} = \frac{ieE_z}{m\omega(1 - k^2/n_0\gamma_S m\omega^2)}, \quad (46.11)$$

where $\gamma_S = n_0^{-1}(\partial n/\partial p)$, according to (31.10) of Part III, is the coefficient of adiabatic compressibility.

Knowing the velocity acquired by the charge in the field of the wave, we can write the mean current density in the form

$$\mathbf{j} = \sum en_0 \mathbf{v}.$$

This summation is carried out over all kinds of particles. Taking into account (46.10) and (46.11), we obtain

$$\mathbf{j}_{\perp} = i \sum \frac{e^2 n_0}{m\omega} \mathbf{E}_{\perp}, \quad (46.12)$$

$$j_{\parallel} = i \sum \frac{\omega e^2 n_0 E_z}{m\omega^2(1 - k^2/n_0\gamma_S m\omega^2)}. \quad (46.13)$$

Making use of the general formula (31.12), we find for the components of the dielectric permeability tensor

$$\epsilon_{\perp} = 1 - \sum \frac{4\pi e^2 n_0}{m\omega^2}, \quad (46.14)$$

$$\epsilon_{\parallel} = 1 - \sum \frac{4\pi e^2 n_0}{m\omega^2(1 - k^2/n_0\gamma_S m\omega^2)}. \quad (46.15)$$

We see that the phenomenon of spatial dispersion takes place in an isotropic plasma: its dielectric properties are described by the tensor ϵ_{ij} which depends on \mathbf{k} and ω . The values ϵ_{\parallel} and ϵ_{\perp} in the direction of propagation and in the perpendicular direction are different.

We can now employ the above formulae to find a concrete law of dispersion in a plasma, making use of the dispersion equations (46.16) and (46.17).

We introduce, first of all, the following important quantities:

$$(\omega_L^{\text{el}})^2 = \frac{4\pi e^2 n_0^{\text{el}}}{m}, \quad (46.16)$$

which is called the plasma frequency or the Langmuir frequency for electrons, and

$$(\omega_L^{\text{ion}})^2 = \frac{4\pi e^2 n_0^{\text{ion}}}{m_{\text{ion}}}, \quad (46.17)$$

which represents the ion plasma frequency and is smaller than the electron plasma frequency by the ratio m/m_{ion} . For a typical concentration of electrons in the plasma ($n_0^{\text{el}} \sim 10^{15} \text{ cm}^{-3}$, $\omega_L^{\text{el}} \sim 6 \times 10^{11} \text{ sec}^{-1}$) the dielectric constant of the plasma always turns out to be smaller than unity. We then have

$$\epsilon_{\perp} = 1 - \sum \frac{(\omega_L)^2}{\omega^2} \approx 1 - \frac{(\omega_L^{\text{el}})^2}{\omega^2}, \quad (46.18)$$

$$\epsilon_{\parallel} = 1 - \frac{(\omega_L^{\text{el}})^2}{\omega^2(1 - k^2/n_0^{\text{el}}\gamma_S m \omega^2)} - \frac{(\omega_L^{\text{ion}})^2}{\omega^2(1 - k^2/n_0^{\text{ion}}\gamma_S m_{\text{ion}} \omega^2)}. \quad (46.19)$$

Taking into account these values, (46.16) gives

$$\frac{\omega^2}{c^2} \left(1 - \frac{(\omega_L^{\text{el}})^2}{\omega^2} \right) - k^2 = 0;$$

hence

$$\omega^2 = (\omega_L^{\text{el}})^2 + c^2 k^2. \quad (46.20)$$

Formula (46.20) defines the law of dispersion of transverse electromagnetic waves in a plasma. For $\omega > \omega_L^{\text{el}}$ there corresponds the propagation of two waves with different polarizations to each frequency.

For $\omega < \omega_L^{\text{el}}$ the values of the wave number k turn out to be imaginary. This means that the waves undergo a damping similar to the ordinary skin effect. The damping coefficient is equal to

$$\kappa = \frac{(\omega_L^{\text{el}})^2}{\omega^2} - 1.$$

Let us now consider a longitudinal wave in the plasma. If the frequency is large in comparison with the ion plasma frequency, $\omega \gg \omega_L^{\text{ion}}$, then from (33.18) and (46.19) we find

$$\epsilon_{\parallel} \approx 1 - \frac{(\omega_L^{\text{el}})^2}{\omega^2(1 - k^2/\gamma_S n_0^{\text{el}} m \omega^2)} = 0.$$

Hence

$$\omega^2 = (\omega_L^{\text{el}})^2 + \frac{k^2}{\gamma_S n_0^{\text{el}} m}. \quad (46.21)$$

From formula (46.21) it follows that the frequency of longitudinal waves is always close to the plasma frequency. They can propagate only if $\omega > \omega_L^{\text{el}}$. Longitudinal waves in a plasma have a simple meaning. In our approximation, where we have assumed that $\omega_L^{\text{ion}} \rightarrow 0$, i.e. that $m_{\text{ion}} \rightarrow \infty$, heavy ions are at rest. In equilibrium the electrons are distributed relative to the ions in the form of electron clouds which have been considered in §41. When the equilibrium is violated the electrons displace themselves relative to the ions at rest and perform oscillations with a plasma frequency ω_L^{el} . The second term is associated with the waves of the adiabatic compression of the gaseous plasma. These waves are similar to sound waves, but, in contrast to the gas of neutral particles, the compression and rarefaction are accompanied by the separation of charges.

The value of γ_S can easily be found from the equation of the adiabatic curve $p/n^k = \text{const}$, if it is assumed that $k = 3$. The value of the exponent of the adiabatic curve corresponds to one degree of freedom, motion in the direction of propagation of the wave. Then $1/\gamma_S n_0 m = 3\overline{v^2}$, so that

$$\omega^2 = (\omega_L^{\text{el}})^2 + 3\overline{v^2} k^2, \quad (46.22)$$

where $(\overline{v^2})^{\frac{1}{2}}$ is the root-mean-square velocity of thermal motion of the particles. As the wavelength decreases the second term of (46.22) increases. How-

ever, for $\overline{v^2}k^2 \sim \omega_L^{\text{el}}$, which, as is easily seen, corresponds to $\lambda \sim \kappa^{-1}$, the phase velocity of the waves turns out to be comparable with the thermal velocity of the electrons. In this case a strong damping of the longitudinal plasma waves occurs. For $\omega < \omega_L^{\text{el}}$, as is seen from (46.21), the values of k turn out to be imaginary, which corresponds to the damping of the waves. The penetration, as is seen from (46.21) turns out to be equal to the Debye length. Thus the plasma oscillations have frequencies which are close to the plasma frequency ω_L^{el} . Waves with all other frequencies either do not penetrate into the plasma or are rapidly damped.

We shall not dwell on the waves corresponding to the oscillations of ions *.

Let us now consider phenomena occurring in a plasma placed in an external magnetic field. In order to avoid complicated calculations, we shall restrict ourselves to the case of waves whose direction of propagation coincides with the direction of the external field \mathbf{H}_0 . Setting $\mu \approx 1$, instead of (46.1) we can write

$$mn_0 \frac{dv}{dt} = en_0 [\mathbf{E} + c^{-1}(\mathbf{v} \times \mathbf{H}_0)] - \nabla p.$$

We assume that the external magnetic field is large compared with the field of the electromagnetic wave.

For what follows we shall need to take the polarization of the wave into account. We shall assume it to be circularly polarized, so that

$$\mathbf{E} = A(\mathbf{e}_1 \pm i\mathbf{e}_2) \exp [i(kz - \omega t)],$$

where the unit vectors \mathbf{e}_1 and \mathbf{e}_2 are directed along the x -axis and y -axis respectively, and the direction of propagation is chosen as the z -axis. Reproducing all the previous calculations, we easily obtain, instead of (46.10),

$$\mathbf{v}_\perp = i \frac{e\mathbf{E}_\perp}{m(\omega \pm \omega_C)}, \quad (46.23)$$

* See, for example, *Voprosy teorii plazmy (Problems of the theory of plasmas)* No. 3, (Moscow, 1963); V.D. Shafranov, *Elektromagnitnye volny v plazme (Electromagnetic waves in plasma)* V.L. Ginzburg, *Propagation of electromagnetic waves in plasma* (North-Holland Publ. Co., Amsterdam, 1961); in which the whole range of problems concerning electromagnetic processes in plasma is discussed in detail.

where ω_C is the cyclotron frequency eH_0/mc . Correspondingly, instead of (46.18) we find

$$\epsilon_{\perp} = 1 - \frac{(\omega_L^{\text{el}})^2}{(\omega \pm \omega_C)^2}. \quad (46.24)$$

We see that the values of the dielectric constant ϵ_{\perp} or the refractive index $n = \epsilon_{\perp}^{1/2}$ turn out to be dependent on the sense of polarization. To the right-handed (+ sign) and left-handed (− sign) polarizations there correspond different values of the refractive index. In other words, n has different values depending on the relation between the direction of rotation of the vector \mathbf{E} in the wave and the motion of the electron in a circular orbit in the magnetic field. This phenomenon is called double refraction. The phenomenon of double refraction is characteristic of anisotropic media (crystals). We see that the plasma assumes anisotropic high-frequency properties in external magnetic fields.

When $\omega \approx \omega_C$ the refractive index for a wave with left-handed polarization becomes very large, which corresponds to the reflection of the wave from the plasma, whereas a wave with right-handed polarization still penetrates into the plasma.

Important Integrals

Stirling's Formula

Calculation of certain integrals

1) $I = \int_{-\infty}^{\infty} e^{-\alpha x^2} dx$ (the Poisson integral),

$$I = 2 \int_0^{\infty} e^{-\alpha x^2} dx = \frac{2}{\sqrt{\alpha}} \int_0^{\infty} e^{-t^2} dt .$$

We have the identity

$$I^2 = \frac{4}{\alpha} \int_0^{\infty} e^{-t^2} dt \int_0^{\infty} e^{-u^2} du = \frac{4}{\alpha} \int_0^{\infty} \int_0^{\infty} e^{-(u^2+t^2)} dt du .$$

Introducing polar coordinates

$$r^2 = u^2 + t^2 ,$$

$$\varphi = \arctan \frac{u}{t} ,$$

$$dt du = r dr d\varphi ,$$

we have

$$I^2 = \frac{4}{\alpha} \int_0^{\pi/2} \int_0^{\infty} e^{-r^2} r \, dr \, d\varphi = \pi/\alpha.$$

hence

$$I = \sqrt{\pi/\alpha}.$$

$$2) \quad I_{2n} = \int_{-\infty}^{\infty} e^{-\alpha x^2} x^{2n} \, dx.$$

Differentiating I with respect to the parameter α , we find

$$I_2 = \int_{-\infty}^{\infty} e^{-\alpha x^2} x^2 \, dx = \frac{1}{2} \sqrt{\pi/\alpha^3},$$

$$I_4 = \int_{-\infty}^{\infty} e^{-\alpha x^2} x^4 \, dx = \frac{3}{4} \sqrt{\pi/\alpha^5},$$

$$I_{2n} = \int_{-\infty}^{\infty} e^{-\alpha x^2} x^{2n} \, dx = \frac{(2n-1)(2n-3) \dots 5 \cdot 3 \cdot 1}{2^n} \sqrt{\pi/\alpha^{2n+1}}.$$

$$3) \quad I_{2n+1} = \int_0^{\infty} e^{-\alpha x^2} x^{2n+1} \, dx,$$

$$I_1 = \int_0^{\infty} e^{-\alpha x^2} x \, dx = 1/2\alpha.$$

Differentiating I_1 with respect to the parameter α , we obtain

$$I_{2n+1} = \int_0^{\infty} e^{-\alpha x^2} x^{2n+1} dx = \frac{n!}{2\alpha^{n+1}}.$$

$$\begin{aligned} 4) \quad \int_0^{\infty} \ln(1 - e^{-x}) dx &= x \ln(1 - e^{-x}) \Big|_0^{\infty} - \int_0^{\infty} \frac{x dx}{e^x - 1} = \\ &= - \int_0^{\infty} \frac{x dx}{e^x - 1}. \end{aligned}$$

Analogously,

$$\begin{aligned} \int_0^{\infty} x^2 \ln(1 - e^{-x}) dx &= \frac{1}{3} x^3 \ln(1 - e^{-x}) \Big|_0^{\infty} - \frac{1}{3} \int_0^{\infty} \frac{x^3 dx}{e^x - 1} = \\ &= -\frac{1}{3} \int_0^{\infty} \frac{x^3 dx}{e^x - 1}. \end{aligned}$$

The calculation of the last integrals is carried out by means of the expansion of the integrand in a power series:

$$\frac{1}{e^x - 1} = \sum_{n=0}^{\infty} e^{-x} e^{-nx} = \sum_{n=0}^{\infty} e^{-(n+1)x}.$$

Hence it follows that

$$\begin{aligned} \int_0^{\infty} \frac{x dx}{e^x - 1} &= \sum_{n=0}^{\infty} \int_0^{\infty} x e^{-(n+1)x} dx = \sum_{n=0}^{\infty} \frac{1}{(n+1)^2} = \frac{\pi^2}{6}, \\ \int_0^{\infty} \frac{x^3 dx}{e^x - 1} &= \sum_{n=0}^{\infty} \int_0^{\infty} x^3 e^{-(n+1)x} dx = \sum_{n=0}^{\infty} \frac{6}{(n+1)^4} = \frac{6\pi^4}{90} = \frac{\pi^4}{15}. \end{aligned}$$

$$\begin{aligned}
 5) \quad I &= \int_{-\infty}^{\infty} x^2 \frac{d}{dx} \frac{1}{e^{-x} + 1} dx = \\
 &= \int_{-\infty}^{\infty} \frac{x^2 e^{-x} dx}{(e^{-x} + 1)^2} = 2 \int_0^{\infty} \frac{x^2 e^{-x} dx}{(e^{-x} + 1)^2}.
 \end{aligned}$$

Expanding the integrand in a power series, we have

$$\begin{aligned}
 I &= 2 \int_0^{\infty} x^2 (e^{-x} - 2e^{-2x} + 3e^{-3x} - \dots) dx = \\
 &= 4 \left(1 - \frac{1}{2^2} + \frac{1}{3^2} - \dots \right) = 4 \cdot \frac{\pi^2}{12} = \frac{\pi^2}{3}.
 \end{aligned}$$

Stirling's formula

Stirling's formula

$$N! \approx N^N e^{-N} (2\pi N)^{\frac{1}{2}}$$

holds for large values of the number N . It is obtained from a simple calculation to an accuracy within the factor $\sqrt{2\pi}$. Namely,

$$\ln N! = \sum_{n=1}^N \ln n,$$

and further, according to the Euler–Maclaurin formula,

$$\sum_{n=1}^N \ln n \approx \int_1^N \ln x dx + \frac{1}{2} \ln x \Big|_1^N + C = N \ln N - N + \frac{1}{2} \ln N + C.$$

A more accurate calculation leads to the value $C = \sqrt{2\pi}$.

SUBJECT INDEX

- Absorption, 538, 543
 - coefficient, 535, 551
- Absorptive capacity, 358
- Acoustic branch, 255
- Activity, 298
 - , grand partition function as a function of, 317
- Addition of probabilities, 23
- Additivity, 100
 - of the entropy, 111
- Adiabatic compressibility, 135, 136
- Adsorbed particles, chemical potential of, 332
 - —, free energy of, 332
- Adsorbent, 331
- Adsorption, gas, 331
 - isotherm, 333
- Alfvén magnetohydrodynamic waves, 620
- Alternating current, 495
 - — machine, 521
 - —, Ohm's law for, 500
- Angular momentum, intrinsic, 470
- Anisotropic media, 417
- Anode region, 596
- Anomalous dispersion, region of, 554
- Antiferromagnet, 417, 466
- Area of quantum state in phase space, 12
- Arithmetical mean value, 25
- Asymmetry of second law and asymmetry of initial condition, 157
- Attenuation of correlation at infinity, 36
- Attraction, internal pressure due to, 228
- Autocorrelation function of the random e.m.f., 526
- Babinet's principle, 564
- Band, energy, 18
- Basic potential, 129
 - thermodynamic equality, 109
 - — inequality, 117
- Binary distribution function, 231
 - — —, determination of, 240
 - — —, energy in terms of, 238
 - — —, equation of state in terms of, 237
- Binomial law, 32
- Biot-Savart law, 462
- Black-body, absolute, 359
 - — radiation, classical theory of, 360
 - — —, free energy of, 370
- Bohr quantum condition, 7, 11
- Boltzmann constant, 118
 - distribution in a plasma, 598
 - formula, 113
- Bose-Einstein distribution, 350
- Boundary conditions, 413
 - — at the interface of two dielectrics, 424
 - —, Dirichlet's, 428
 - —, Neumann's, 428
 - — on current, 454
 - — on electric field, 414
 - — on magnetic field, 415
- Bound charges, current density of, 408
- Bound energy, 126
- Box, three-dimensional, particle in, 12
- Bragg's conditions, 569
- Bremsstrahlung, 585
- Brownian force, 278
 - motion, 276
 - —, experiments on, 281
 - —, quantitative theory of, 278
 - particle, mean square displacement of, 279

- Brownian particles, number of, in the viewing field, 282
- Canonical distribution, 81
 - -, properties of, 84
- Capacitance, 426
 - -, electrostatic, 458
- Carnot cycle, 121
- Cartesian coordinates, Laplace's equation in, 428
- Cathode region, 596
- Causality, 540
- Cells, method of, in phase space, 346
- Central limit theorem, 31
- Cerenkov counter, 592
 - loss, 589, 590
 - Vavilov radiation, 585, 592
- Characteristic function of probability distribution, 34
 - temperature for rotation, 209
 - - for vibration, 206
 - - of polyatomic molecules, 216
- Charge conservation law, 398
 - density in a plasma, 599, 601
 - -, mean, 403
 - -, free, 406
- Chemical compounds, thermodynamic potentials of, 144
 - equilibrium, 335
 - - condition, 337
 - potential, 301, 302
 - - and phase transition, 309
 - - of a crystal, 304
 - - of adsorbed particles, 332
 - - of an ideal gas, 303
 - - of photon gas, 368
 - - of the electron gas, 390
- Clapeyron-Clausius equation, 309
- Classical approximation, Gibbs distribution in, 87
 - mechanics, transition from quantum mechanics to, 9
 - physics, magnetic moment in, 468
 - statistics, 86
 - -, applicability of, 353
- Closed system, 111
 - -, entropy for irreversible process in, 122
 - -, entropy for reversible process in, 122
 - -, fluctuations in, 271
- Coefficient of mutual inductance, 501
 - - - - for non-linear conductors, 510
 - - - - of linear conductors, 503
 - - - - of two solenoids, 504
 - - self-inductance, 501, 507
 - - - for non-linear conductors, 510
 - - - of toroidal solenoid, 508
 - - - of two coaxial cylinders, 508
- Collision, energy conservation law for, 45
 - , number of, 62
 - of molecules, 59
 - , particle, 44
- Complementary screen, 564
- Compressibility, adiabatic, 135, 136
 - -, coefficient of, 624
 - , isothermal, 135
- Condensation, 321
- Conditions of equilibrium between two phases, 305
- Conduction current density, 405
- Conductivity, electrical, 405, 452
 - -, tensor, 549
 - , Hall, 609
 - in a plasma, 606
- Conductor, 400
 - , electromagnetic energy of, 435
 - , electrostatic field in, 424
 - , ideal, 577
 - , interaction between, 519
 - , linear, 455
 - , moving, 512, 517
 - -, force on, 517
 - -, energy of system of, 512
 - , surface charge density for, 425
 - , surface potential of, 426
- Configuration integral, 222
 - , probability of, 230
 - space, 5
- Constitutive equations, 410
- Container wall, 48
- Coordinates, normal, 254
- Correlation function, 35, 524
 - - at infinity, attenuation of, 36

- - in a plasma, 602
- - method, 230
- Critical angle for total reflection, 575
 - magnetic field of superconductor, 489
 - point, 321
 - -, conditions for, 322
 - -, position of, 288
 - -, properties at, 324
- Cross section, 61
- Crystal, atom distance in, 242
 - , chemical potential of, 304
 - , entropy of, 264
 - , equation of state of, 265
 - , free energy of, 264
 - , heat capacity in, quantum-mechanical, 244
 - , - - of, 243, 263, 267
 - , mean energy of, 243
 - , one-dimensional, dispersion law in, 247
 - , -, kinetic energy of, 252
 - , -, thermal motion in, 245
 - , -, total energy of, 253
 - , partition function of, 260
 - , thermodynamic functions of, 263
 - , three-dimensional, energy of, 256
 - with different masses, 254
- Curie point, 478
 - -, entropy at, 484
 - -, heat capacity at, 484
 - -, magnetic properties near, 483
 - Weiss law, 478
- Current, alternating, 495
 - , boundary conditions on, 454
 - , correlation function for, 525
 - density of bound charges, 408
 - -, conduction, 405
 - -, mean, 403, 409
 - , direct, magnetic fields of, 460
 - , energy of, 457
 - , fluctuation, 523
 - loops, moving, 496
 - , polarization, 407
 - , random, 524
 - sources, 452
 - , surface, in superconductor, 488
- Cyclic process, 98
- Cylinder in electrostatic field, 437
 - , two coaxial, coefficient of self-inductance of, 508
- Damping coefficient for plasma waves, 626
- Debye characteristic temperature, 261, 266
 - cut-off, 259
 - length, 601
- Degeneracy, 13
- Degenerate gas, 353
 - -, condition for, 355
 - quantum state, 13
 - state, 190
- Density, fluctuation of, 286
 - , spectral, 37
- Determination of binary distribution function, 240
- Deviation, mean square, 28
 - , - -, in normal distribution, 33
- Diamagnet, 412, 465, 467
- Diamagnetic susceptibility, 469
- Diatomic gas, 179
 - -, heat capacity for, 187
 - -, thermodynamic functions of, 200
 - molecule, 195
 - -, basic quantities of, 198
 - -, energy levels of, 195
 - -, Gibbs distribution for, 181
 - -, rotational motion of, 199
 - -, vibrational motion of, 197
- Dielectric, 400
 - constant, 411, 449
 - -, longitudinal, 546
 - -, perpendicular, 546
 - , free energy of, 443
 - in electrostatic field, 437
 - permeability tensor, 624
 - permittivity, 411, 541
 - -, complex, 535
 - susceptibility, 407, 445
 - , thermodynamic potential of, 444
 - , two, boundary conditions at the interface of, 424
- Diffacted radiation, intensity of, 567
- Diffraction, 560
 - , Fraunhofer, 560
 - , X-ray, 568

- Diffuseness, zone of, 386
- , —, electrons in, 392
- Diffusion coefficient, 279
- Dipole moment, 445
 - , —, induced, 450
 - , —, of a gas, 447
 - , —, of a sphere, 440
 - , —, of a dielectric, 442
 - , —, of molecules, 449
- Dirichlet's boundary conditions, 428
- Discrete energy levels, 8
- Disorder, molecular, 42
- Dispersion, 249, 543
 - , anomalous, region of, 554
 - , frequency, 419
 - , law in conducting medium, 536
 - , —, in one-dimensional crystal, 247
 - , —, in quantum mechanics, 552
 - , —, in rarefied gas, 551
 - , of light, 549
 - , relation, 540, 543, 547
 - , spatial, 419, 544
 - , time, 419
- Displacement law, 365
 - , mean square, of Brownian particle, 279
- Dissociation, degree of, 341
 - , law of mass action for, 340
 - , of atoms, thermal, 340
- Distance between atoms in crystals, 242
 - , between molecules, mean, 222
- Distribution, Bose-Einstein, 350
 - , canonical, 81
 - , —, properties of, 84
 - , Fermi-Dirac, 351, 382, 387
 - , function, binary, 231
 - , —, —, determination of, 240
 - , —, —, energy in terms of, 238
 - , —, —, equation of state in terms of, 237
 - , —, differential equation for, 232
 - , —, in photon gas, 368
 - , —, of m th order, 231
 - , —, ordinary, 231
 - , Gibbs, 75, 81, 232
 - , —, for diatomic molecules, 181
 - , —, for ideal gas, 167
 - , —, for one molecule, 89
 - , —, in the classical approximation, 87
 - , —, maximum in, 85, 93
 - , —, of monatomic gas, 91
 - , —, properties of, 84
 - , —, sharpness of maximum in, 94
 - , —, grand canonical, 297
 - , microcanonical, 75
 - , modulus, 82
 - , statistical, 68
 - , —, of system with variable number of particles, 295
- Double refraction, 628
- Dulong-Petit law, 243
- Dynamical law, 6
- Easy magnetization, direction of, 478
- Effective electrons, 392
- Efficiency, 121
 - , maximum, 122
- Eikonal, 556
 - , equation, 557
- Electric displacement, 410
 - , field, boundary conditions on, 414
 - , —, of a superconductor, 488
 - , —, static, plasma in, 605
 - , susceptibility of a polar gas, 448
- Electro-caloric effect, 444
- Electrodynamics, macroscopic, 397
- Electromagnetic field, penetration distance of, 575
 - , potentials, 412
 - , —, equations for, 412
 - , radiation, energy of, 369
 - , waves, group velocity of, 553
 - , —, in a medium, 533
- Electromotive force (*see also* e.m.f.)
 - , —, impressed, 453, 456
- Electron, energy of, at absolute zero, 383
 - , gas, 355
 - , —, at absolute zero, 381
 - , —, at low temperatures, 385
 - , —, chemical potential of, 390
 - , —, energy of, 387
 - , —, ground state of, 383
 - , —, heat capacity of, 384, 390
 - , —, interaction energy in, 385
 - , —, pressure of, 384

- -, thermal excitation at low temperature in, 386
- in the zone of diffuseness, 392
- Electrostatic capacitance, 458
- field, energy relations in, 435
- in a medium, 422
- in conductors, 424
- induction, 401
- Electrostatics, direct problem of, 427
- Electrostriction, 444
- Elementary excitations, 372
- energy spectrum, 374
- number of, 377
- Elementary quantum states, equal probability of, 72
- E.m.f. (*see also* electromotive force)
- autocorrelation function of the random, 526
- random fluctuating, 523
- Emitted energy, 357
- Energy change in quasi-static process, 105
- conservation law, 420
- for collision, 45
- density of radiation, 357, 359, 364
- distribution, 53
- electromagnetic, of conductor, 435
- equipartition of, 57
- in terms of binary distribution function, 238
- internal, 98
- kinetic, mean, 56
- of one-dimensional crystal, 252
- level, 6
- distribution of, in a solid body, 265
- of diatomic molecule, 195
- mean, of a crystal, 243
- of a polyatomic molecule, 186
- of rotation, 211
- of triatomic molecules, 185
- of current, 457
- electron gas, 387
- electrons at absolute zero, 383
- ideal gas, 168
- magnetic field, 505
- plasma, 604
- quasi-independent systems, 67
- subsystem, 68
- surface, 326
- system, 71
- system in a magnetic field, 470
- system of moving conductors, 512
- three-dimensional crystal, 256
- pressure versus, 58
- relations in the electrostatic field, 435
- total, of one-dimensional crystal, 253
- vibrational, mean, 203
- Ensemble, statistical, 22
- Enthalpy, 129
- determination of, 142
- Entropy, 108
- additivity of, 111
- at Curie point, 484
- at low temperature, 147
- change, 114
- in open system, 117
- determination of, 143
- for irreversible process in closed system, 123
- for reversible process in closed system, 122
- in phenomenological thermodynamics, 122
- of crystal, 264
- an ideal gas, 169
- a quasi-closed system, 114
- HCl, 150
- liquid helium II, 378
- mixing, 171
- non-equilibrium system, 112
- normal state, 490
- radiation, 370
- superconducting state, 490
- surface, 326
- properties of, 110
- rotational, 212
- vibrational, 207
- Equal probability of elementary quantum states, 72
- Equation of continuity, 406
- state, 50, 225
- in terms of the binary distribution function, 237
- of a crystal, 265
- of an ideal gas, 169

- Equilibrium, approach to, 114
 - between two phases, conditions of, 305
 - , chemical, 335
 - , -, condition, 337
 - curve, equation of, 313
 - -, phase, 307
 - -, -, differential equation of, 308
- Equilibrium, phase, between drops and vapour, 329
 - plasma, 597
 - radiation, 356
 - , state of, 1
 - state of a reacting system, 336
- Equipartition, law of, 186
 - of energy, 57
- Equiphase surface, 558
- Ergodic hypothesis, 22, 72
 - systems, 72
- Euler-Maclaurin formula, 632
- Exclusion principle, Pauli, 17
- Expansion, thermal, at zero temperature, 151
 - , -, coefficient of, 135
- Experimental determination of the mean free path, 64
 - verification of the Maxwell distribution, 54
- External parameters, 101
- Faraday's law of induction, 498
- Fermi-Dirac distribution, 351, 382, 387
 - energy, 384
- Ferroelectrics, 417
- Ferromagnet, 417, 465, 476
 - , magnetic susceptibility of, 478
 - , metastable states in, 481
 - , thermodynamic potential of, 479
- Fluctuation, 25, 154
 - and thermodynamics, 278
 - , change in potential energy in, 274
 - currents, 523
 - dissipation theorem, 526
 - in a closed system, 271
 - in a quasi-closed system, 272
 - , intensity of, 276
 - of density, 286
 - - number of particles, 287
 - - temperature, 288
 - - volume at a constant temperature, 284
 - , perpetual motion machine and, 161
 - , probability of, 274
 - , relative, 28
- Forbidden states, 73
- Force, generalized, 101
- Fraunhofer diffraction, 560
- Free charges, 406
 - energy, 126, 137
 - - as criterion of reversibility, 127
 - -, electric, 421
 - - of a crystal, 264
 - - - adsorbed particles, 332
 - - - an ideal gas, 168
 - - - black-body radiation, 370
 - - - dielectric, 443
 - - - liquid helium II, 376
 - - - magnetic field, 465
 - - - plasma, 604
 - - per unit surface, 326
 - -, rotational, 212
 - -, vibrational, 207
 - path, mean, 62
 - -, -, experimental determination of, 64
 - -, probability distribution of, 62
- Frequency dispersion, 419
- Fresnel formulae, 574
- Gas, density of, in gravitational field, 176
 - , weight of, 177
- Gaseous discharge, 595
 - -, non-self-maintained, 595
 - -, self-maintained, 595
- Gaussian distribution, 31, 275
- Geometrical optics, 555
 - -, applicability of, 559
- Gibbs distribution, 75, 81, 232
 - - for diatomic molecules, 181
 - - for ideal gas, 167
 - - for one molecule, 89
 - - in the classical approximation, 87
 - -, maximum in, 85, 93
 - - of monatomic gas, 91
 - -, properties of, 84

- -, sharpness of maximum in, 94
- Helmholtz equations, 132
- thermodynamic potential, 109, 126, 138, 299
 - - - of magnetic field, 465
- Glow discharge, 595
- Grand canonical distribution, 297
 - partition function, 298
 - - - and phase transition, 317
 - - - as a function of the activity, 317
 - - -, zeros of, 318
- Gravitational field, density of gas in, 176
 - -, heat capacity of gas in, 177
- Ground state of electron gas, 383
- Group velocity in a wave guide, 583
 - - of electromagnetic waves, 553

Hagen-Rubens formula, 576

Hall conductivity, 609

- current, 609

Hamilton's equations, 3

Harmonic oscillator in quantum mechanics, 12

- -, linear, 4

Heat capacity, 130, 136, 187

- - at Curie point, 484

- - at low temperature, 147, 245

- - due to internal motion, 268

- - in crystal, quantum-mechanical, 244

- - of crystal, 243, 263, 267

- - of diatomic gas, 187

- - of electron gas, 384, 390

- - of gas in gravitational field, 177

- - of ideal gas, 168

- - of lattice, 391

- - of liquid helium II, 378

- - of monatomic gas, 170

- - of normal state, 491

- - of polyatomic gas, 189

- - of superconducting state, 491

- - of system with two levels, 192

- - of unit surface, 327

- -, rotational, 211

- -, vibrational, 203

- engine, 118

- - of the first kind, 119

- - of the second kind, 124

- -, reduced, 122

Height distribution of molecules possessing different masses, 179

Helium II, liquid, entropy of, 378

- -, -, free energy of, 376

- -, -, heat capacity of, 378

- -, -, properties of, 372

- -, -, thermal conductivity of, 379

- -, -, two-fluid model for, 381

- -, -, viscosity of, 379

Histogram, 54

Hysteresis cycle, 477

Ideal conductor, 577

- gas, 39, 164

- -, chemical potential of, 303

- -, energy of, 168

- -, entropy of, 169

- -, equation of state of, 169

- -, free energy of, 168

- -, Gibbs distribution for, 167

- -, heat capacity of, 168

- - in uniform field, 174

- - of particles with spin, 75

- - plasma, 597

- -, thermodynamic potential of, 170

- monatomic gas, partition function of, 164

Identity of particles, 16, 166, 344

Images, method of, in electrostatics, 430

Impedance, 516

Impressed electromotive force, 453, 456

Induced dipole moment, 450

- magnetic moment, 467, 469

Inductance, mutual, coefficient of, 501

- -, -, for non-linear conductors, 510

- -, -, of linear conductors, 503

- -, -, of two solenoids, 504

- -, self-, coefficient of, 501, 507

- -, -, for non-linear conductors, 510

- -, -, of toroidal solenoid, 508

- -, -, of two coaxial cylinders, 508

Induction, electrostatic, 401

- , Faraday's law of, 498

Inertia, moment of, 181

Initial condition, 40, 71

- Initial condition, asymmetry of second law and asymmetry of, 157
- Intensity of fluctuations, 276
- Interacting molecules, partition function for, 225
- Interaction between conductors, 519
 - quasi-independent systems, 67
 - subsystem and reservoir, 77
 - subsystems, 68
 - energy in electron gas, 385
 - , interatomic, 194
 - , intermolecular, 39, 220
 - , pair-, 223
 - , potential energy of, 220
 - , radius of, 221
- Interface between two phases, 305
 - , curvature of, 328
- Internal energy and mean energy, 99
 - motion, heat capacity due to, 268
 - , partition function for, 201
- Inversion, method of, in electrostatics, 434
 - point, 146
- Irreversibility, criterion of, 115, 159
 - , macroscopic, microscopic reversibility versus, 155
- Irreversible process in closed system, entropy for, 123
- Isothermal compressibility, 135
- Isotope effect, 491

- Joule heat, 420
 - Thompson coefficient, 145
 - process, 144

- Kirchhoff theorem, 357
- Kramers—Kronig formulae, 540, 543

- Langevin formula, 448
- Langmuir frequency, 625
- Laplace equations in Cartesian coordinates, 428
 - formula, 328
 - pressure, 328
- Latent heat in phase transition, 307
 - of phase transformation, 312
 - of superconducting transition, 490
- Lattice, heat capacity of, 391
- Laue formula, 569

- Law of mass action, 338
 - — — for dissociation, 340
 - — —, statistical meaning of, 338
- Light, dispersion of, 549
 - ray direction, 558
 - wave, monochromatic, 556
- Linear conductor, 500
 - , magnetic field of current in, 462
- Liquids, statistical theory of, 239
- Longitudinal dielectric constant, 546
 - wave, 257, 548
 - in wave guide, 580
 - , number of, 258
- Lorentz condition, 412

- Macroscopic electrodynamics, 397
 - system, 1
- Magnetic domains, 482
 - field, 410
 - , boundary conditions on, 415
 - , energy of, 505
 - , energy of system in, 470
 - , free energy of, 465
 - , Gibbs thermodynamical potential of, 465
 - in moving plasma, 613
 - of current in linear conductor, 462
 - of direct currents, 460
 - of superconductor, critical, 489
 - , plasma in, 607
 - induction in superconductor, 487
 - isolation, 610
 - moment, 464, 467
 - , induced, 467, 469
 - in classical physics, 468
 - , mean, 409, 468, 473
 - permeability, 411
 - pressure, 611
 - properties near Curie point, 483
 - susceptibility, 411, 465
 - of ferromagnet, 478
- Magnetization, easy, directions of, 478
 - , residual, 477
 - , spontaneous, 477
 - , work of, 520
- Magnetohydrodynamic waves, Alfvén, 620
 - , velocity of propagation of, 621

- Mass action, law of, 338
 - — —, for dissociation, 340
 - — —, statistical meaning of, 338
 - , reduced, 60
- Maximum in Gibbs distribution, 85, 93
 - — —, sharpness of, 94
 - term, method of, 86
- Maxwell-Boltzmann distribution, 175
 - distribution, 44, 47, 175, 345
 - —, experimental verification of, 54
 - equations, 410
 - —, applicability of, 416
 - — for quasistationary fields, 499
 - —, integral form of, 412
 - Lorentz equations, 398
 - particles, statistical weight for, 346
 - relations, 131
- Mean absolute velocity, 56
 - energy, internal energy and, 99
 - free path, 62
 - — —, experimental determination of, 64
 - kinetic energy, 56
 - square deviation, 28
 - — — in normal distribution, 33
 - value, 25
 - —, arithmetical, 25
 - — for a physical quantity, 82
 - —, statistical, 26
- Measuring devices, sensitivity of, 290
- Metamagnetics, 466
- Metastable states in ferromagnet, 481
- Melting point, 324
- Microcanonical distribution, 75
- Microparticles, 1
- Microscopic reversibility, 74
- Mirror, sensitivity of suspended small, 291
- Mixing, entropy of, 171
- Mobility, 280
- Molar fractions, 339
- Molecular disorder, 42
- Molecules, collisions of, 59
- Moment of inertia, 181
 - of n th order, 34
- Momentum distribution, 53
 - of particle, generalized, 466
 - space, 5
- Monatomic gas, Gibbs distribution of, 91
- Monochromatic light wave, 556
 - plane waves, 534
- Multiplication of probabilities, 24
- Nernst heat theorem, 148
- Neumann's boundary conditions, 428
- Noise, 523
- Non-cyclic process, maximum work in, 124
- Non-degenerate gas, 353
- Non-equilibrium properties, 2
 - system at zero temperature, 151
 - —, entropy of, 112
- Non-local connection, 544
- Normal coordinates, 254
 - distribution, 31
 - —, mean square deviation in, 33
 - liquid, 381
 - oscillation of a crystal, 254
 - state, entropy of, 490
 - —, heat capacity of, 491
 - —, thermodynamic potential of, 490
- Normalization of probability, 25
- N -particle system, relative fluctuation in, 28
- Number density of states, 15
 - of particles, fluctuation of, 287
 - — —, variable, subsystems with, 294
 - — —, —, statistical distribution of system with, 295
 - — states, 14
 - — —, Taylor series for, 79
 - — — with given energy, 15
- Nyquist formulae, 527
- Ohm's law, 452
 - — for an alternating current, 500
 - — in differential form, 405
 - — in generalized form, 453
- Open system, entropy change in, 117
- Optical branch, 256
 - density, 573
 - path length, 556
- Optics, geometrical, 555
 - —, applicability of, 559
- Ordinary distribution function, 231
- Orthohydrogen, 213

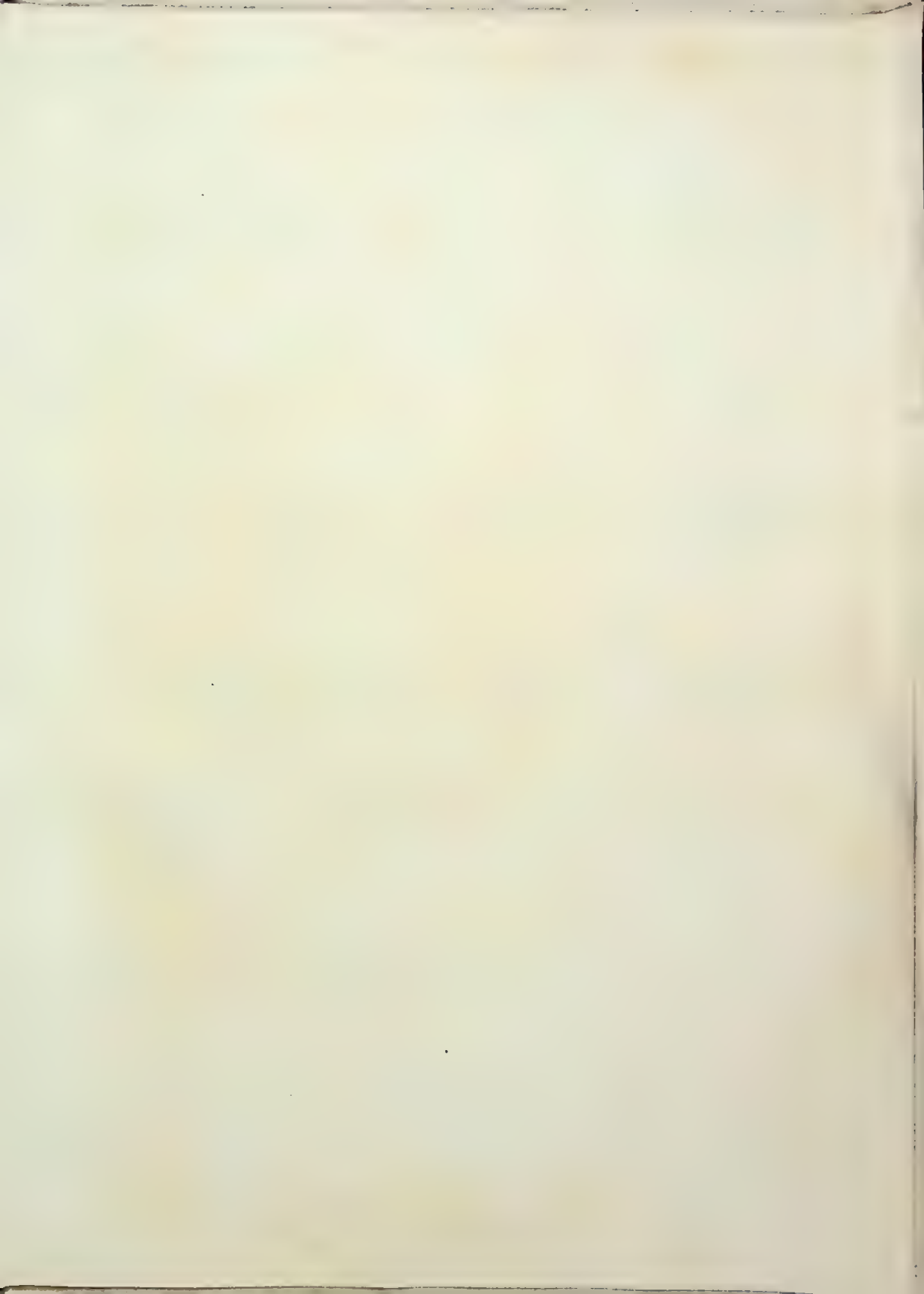
- Pair-interaction, 223
- Parahydrogen, 213
- Paramagnet, 411, 465, 468
 - , partition function of, 472
- Paramagnetic susceptibility, 471, 473
 - , verification of the theory of, 475
- Partition function, 81, 86, 137
 - for interacting molecules, 225
 - for the internal motion, 201
 - for the translated motion, 201
 - , grand, 298
 - in terms of the pressure, 140
 - of crystal, 260
 - of ideal monatomic gas, 164
 - of polyatomic molecule, 213
 - of system in magnetic field, 467
 - , rotational, 208
 - , —, of polyatomic molecule, 215
 - , vibrational, 202
 - , —, of polyatomic molecule, 216
- Pauli exclusion principle, 17
- Penetration distance of electromagnetic field, 575
- Periodicity, conditions of, 249
- Permeability, 544
 - , magnetic, 411
 - , —, of superconductor, 489
 - of plasma, 607
 - tensor, dielectric, 624
- Permittivity, dielectric, 411, 541
 - , —, complex, 535
- Perpendicular dielectric constant, 546
- Perpetual motion machine and fluctuations, 161
 - — — of the second kind, 119, 155
- Pinch effect, 612
- Phase equilibrium between drops and vapour, 329
 - — curve, 307
 - — —, differential equation of, 308
 - of substance, 305
 - space, 3
 - —, area of quantum state in, 12
 - —, volume element in, 5
 - — volume of quantum state, 13
 - trajectory, 4
 - transformation, latent heat of, 312
 - transition, chemical potentials and, 309
 - —, grand partition function and, 317
 - —, latent heat in, 307
 - — of the first kind, 316
 - — of the second kind, 316, 320, 372
 - —, superconducting, 489
 - , two, conditions of equilibrium between, 305
 - , —, interface between, 305
 - velocity in a wave-guide, 583
 - volume, 14
- Phenomenological thermodynamics, entropy in, 122
- Phonons, 373
- Photon, angular momentum of, 366
 - , energy of, 366
 - gas, 355, 366
 - —, chemical potential of, 368
 - —, distribution function in, 368
 - , momentum of, 366
- Physical quantity, mean value for, 82
- Planck constant, 7
 - formula, 363, 369
- Plasma, 596
 - , Boltzmann distribution in, 598
 - , charge density in, 599, 601
 - , conductivity in, 606
 - , correlation functions in, 602
 - , energy of, 604
 - , equilibrium, 597
 - , —, thermodynamics of, 604
 - , free energy of, 604
 - frequency, 625
 - , ideal gas, 597
 - in magnetic field, 607
 - in static electric field, 605
 - -like media, 545
 - , motion of, equations of, 610
 - , moving, magnetic field in, 613
 - , permeability of, 607
 - , pressure in, 622
 - , pressure of, 604
 - , screening in, 602
 - two-fluid model, 622
- Poisson-Boltzmann equation, 600
 - formula, 32
 - integral, 629
- Polar gas, electric susceptibility of, 448
 - liquids, 450

- molecule, 446
- Polarization, 401, 442
 - coefficient, 407
 - current, 407
 - , dielectric, of solid, 450
 - energy loss, 585
 - loss, 589, 590
 - vector, 400
 - waves, 548
- Polyatomic gas, heat capacity for, 189
 - molecule, characteristic temperatures of, 216
 - —, mean energy of, 186
 - —, partition function of, 213
 - —, rotational motion of, 214
 - —, rotational partition function of, 215
 - —, vibrational motion of, 215
 - —, vibrational partition function of, 216
- Potential, electromagnetic, 412
 - in a plasma, 599
 - , thermodynamic, 109, 127
 - —, Gibbs, 109, 126, 138, 299
- Poynting vector, 420
- Pressure, 48, 102
 - in plasma, 622
 - , internal, due to attraction, 228
 - , magnetic, 611
 - of electron gas, 384
 - of plasma, 604
 - , radiation, 370
 - , thermal coefficient of, 135
 - versus energy, 58
- Primitive lattice, 568
- Probabilities, addition of, 23
 - , multiplication of, 24
- Probability density, 22
 - — for the velocity, 52
 - distribution, characteristic function of, 34
 - — of free paths, 62
 - , equal a priori, 73
 - for the subsystem in the state with energy ϵ , 80
 - of a fluctuation, 274
 - of the i th state, 20
 - , normalization of, 25
 - with respect to the ensemble, 22
- Process, isothermal-isobaric, 127
 - , isothermal-isovolumic, 127
- Quantization condition, 10
- Quantized quantity, 6
- Quantum mechanics, harmonic oscillator
 - in, 12
 - —, transition to classical mechanics from, 9
 - number, 7
 - particles, statistical weight for, 349
 - state, area of, in phase space, 12
 - —, degenerate, 13
 - —, phase space volume of, 13
- Quasi-classical approximation, 9
- Quasi-closed system, entropy of, 114
 - —, fluctuations in, 272
- Quasi-independent systems, 66
 - —, energy of, 67
 - —, interaction between, 67
- Quasi-static process, 103
 - —, demands on, 106
 - —, energy change in, 105
- Quasistationary field, 494
 - —, equations of, 495
 - —, Maxwell equations for, 499
- Quasistationarity, first condition of, 493
 - , second condition of, 494
 - , third condition of, 494
- Radiation, black-body, free energy of, 370
 - by fast particles, 585
 - , energy density of, 357, 359, 364
 - , entropy of, 370
 - pressure, 370
 - , thermodynamic potential of, 370
 - , transitional, 593
- Radiative capacity, 358
 - — of an absolute black body, 359
- Radius of interaction, 221
- Raman scattering, 555
- Random function, 35
- Ray direction, light, 558
- Rayleigh-Jeans law, 361
- Reacting system, equilibrium states of, 336

- Recovery time, 160
- Reduced heat, 122
 - mass, 60
- Reflected wave, 571
 - -, amplitude of, 574
- Reflection, angle of, 572
 - internal, total, 576
 - total, critical angle for, 575
- Reflectivity, 574
- Refracted wave, 571
 - -, amplitude of, 574
- Refraction, angle of, 572
- Refractive index, 535, 537, 551, 572
 - - in quantum mechanics, 552
- Relative motion, mean velocity of, 61
 - velocity distribution, 60
- Representative point, 3
- Reservoir, 70
 - interaction between subsystem and, 77
 - subsystem in, 77
- Resistance, 456, 526
- Retardation, 493
- Reversibility, criterion of, 115, 159
 - free energy as criterion of, 127
 - microscopic, 74
 - -, versus macroscopic irreversibility, 155
 - thermodynamic potential as criterion for, 128
- reversible process, 103
 - in a closed system, entropy for, 122
- RLC*-series circuit, 515
- Root-mean-square velocity, 59
- Rotational constant, 199
 - entropy, 212
 - free energy, 212
 - heat capacity, 211
 - motion of diatomic molecule, 199
 - of polyatomic molecule, 214
 - partition function, 208
 - - of polyatomic molecule, 215
- Rotation, characteristic temperature for, 209
 - energy of, 180, 181
 - mean energy of, 211
- Rotor, 521
- Screening in a plasma, 602
- Second law, asymmetry of, and asymmetry of initial condition, 157
- Selection rules, 72
- Sensitivity of gas thermometer, 292
 - of spring-balance, 292
 - of suspended small mirror, 291
 - of measuring devices, 290
- Sites, 331
- Skewness, 34
- Skin depth, 530
 - effect, 528
- Solenoid, toroidal, coefficient of self-inductance of, 508
 - two, coefficient of mutual inductance of, 504
- Spatial dispersion, 419, 544
- Spectral density, 37
 - function, 259
- Spectrum, 6
- Sphere, dipole moment of, 440
 - in electrostatic field, 438
- Spin, 16, 470
 - orientation of, 75
 - particles with, ideal gas of, 75
- Spring-balance, sensitivity of, 292
- Stability, condition of, 285, 289
- State of a system, 70
 - degenerate, 13, 190
 - with energy ϵ , probability for subsystem in, 80
- Statistical distribution, 68
 - of system with variable number of particles, 295
 - ensemble, 22
 - independence, 24
 - integral, 88
 - law, 2, 40
 - meaning of law of mass action, 338
 - mean value, 26
 - sum, 81
 - temperature, 81, 82, 117
 - theory of liquids, 239
 - thermodynamics, 99
 - weight, 13, 190
 - for Maxwell particles, 346
 - for quantum particles, 349
- Statistics, classical, 86

- Stator, 521
- Stefan-Boltzmann law, 370
- Stirling's formula, 632
- Stochastic process, 35
- Stokes' formula, 280
- Sublimation curve, equation of, 313
- Subsystem, 68
 - and reservoir, interaction between, 77
- Sum over states, 81
- Superconducting phase transition, 489
 - state, entropy of, 490
 - —, heat capacity of, 491
 - —, thermodynamic potential of, 490
 - transition, latent heat of, 490
 - — temperature, 486
- Superconductivity, 486
- Superconductor, 417
 - , critical magnetic field of, 489
 - , electric field of, 488
 - , magnetic induction in, 487
 - , magnetic permeability of, 489
 - , surface current in, 488
 - , thermodynamic potential of, 489
- Superfluidity, 379
- Superfluid liquid, 381
- Superposition approximation, 241
- Surface charge, 402, 403
 - — density for conductors, 425
 - , energy of, 326
 - , entropy of, 326
 - , free energy per unit, 326
 - , heat capacity of unit, 327
 - potential of conductor, 426
 - tension, 326
- Susceptibility, diamagnetic, 469
 - , dielectric, 407, 445
 - , magnetic, 411, 465
 - —, of a ferromagnet, 478
 - , paramagnetic, 471, 473
 - —, verification of theory of, 475
- System with two levels, 190
 - — — —, heat capacity of, 192
- Taylor series for number of states, 79
- Temperature, absolute, 50
 - —, scale of, 141
 - coefficient of equilibrium pressure, 310
- , different, systems with, 83, 115
- , fluctuations of, 288
- , statistical, 81, 82, 117
- , zero, non-equilibrium systems at, 151
- TE-wave, 581
- Thermal coefficient of pressure, 135
 - conductivity of liquid helium II, 379
 - death, 162
 - excitation at low temperature in electron gas, 386
 - expansion at zero temperature, 151
 - —, coefficient of, 135
 - motion, 9
 - — in a one-dimensional crystal, 245
- Thermally isolated systems, 116
- Thermodynamical law, 2
- Thermodynamic equality, basic, 109
 - — for systems with variable number of particles, 301
 - functions of a crystal, 263
 - — of diatomic gases, 200
 - inequality, basic, 117
 - potential as criterion for reversibility, 128
 - —, Gibbs, 109, 126, 138, 299
 - —, —, of magnetic field, 465
 - — of a ferromagnet, 479
 - — of an ideal gas, 170
 - — of a superconductor, 489
 - — of chemical compounds, 144
 - — of dielectric, 444
 - — of radiation, 370
 - — of the normal state, 490
 - — of the superconducting state, 490
 - —, physical meaning of, 299
 - quantities, determination of, 137, 141
 - —, transformation of, 132
- Thermodynamics, 99
 - , first law of, 98, 104
 - , fluctuations and, 278
 - of equilibrium plasma, 604
 - , second law of, 98
 - , statistical, 99
 - , third law of, 147, 151
- Thermomechanical effect, 381
- Thermometer, gas, sensitivity of, 292
- Third law of thermodynamics, 147, 151
- Three-body problem, 40

- Throttling, 144
- Time dispersion, 419
- TM-wave, 581
- Townsend discharge, 595
- Trajectory of individual molecule, 41
- Transformation of thermodynamic quantities, 132
- Transition from quantum mechanics to classical mechanics, 9
- Transitional radiation, 593
- Transition temperature, superconducting, 486
- Transcritical state, 322
- Translated motion, partition function for, 201
- Translation, energy of, 180
- Transmissivity, 574
- Transparent medium, 537
- Transverse electric wave, 581
 - magnetic wave, 581
 - wave, 257, 548
 - — in wave guide, 580
 - —, number of, 258
- Triatomic molecules, 185
 - —, mean energy of, 185
- Two-fluid model for liquid helium II, 381
 - — for plasma, 622
- Ultraviolet catastrophe, 362
- Uniform density, 42
 - field, ideal gas in, 174
- Vapour pressure, 314
 - — constant, 313
- Velocity, absolute, mean, 56
 - distribution function, 43
 - —, uniform, 42
 - , mean, 55
 - , most probable, 58
 - of a wave, 248
 - of relative motion, mean, 61
 - , probability density for, 52
 - , root-mean-square, 59
- Vibration, characteristic temperature for, 206
 - , energy of, 180, 183
- Vibrational energy, mean, 203
 - entropy, 207
 - free energy, 207
 - heat capacity, 203
 - motion of diatomic molecule, 197
 - — of polyatomic molecule, 215
 - partition function, 202
 - — — of polyatomic molecule, 216
- Viewing field, number of Brownian particles in, 282
- Viscosity of liquid helium II, 379
- Volume, fluctuations of, at constant temperature, 284
- Waals, van der — equation, 225, 226, 229
- Wave, electromagnetic, in a medium, 533
 - guide, 577
 - —, group velocity in, 583
 - —, phase velocity in, 583
 - , plane, monochromatic, 534
 - , velocity of, 248
- Weight of a gas, 177
- Wiener-Khinchin theorem, 38, 526
- Wien's law, 364
- Work, 102
 - , maximum, in non-cyclic process, 124
 - , useful, 124
- X-ray diffraction, 568
- Zero, absolute, unattainability of, 149
 - point energy, 204
- Zeros of grand partition function, 318



Translated from the Russian by
S. Subotić, Belgrade

Translation edited by
J. Schneps, Tufts University, Medford, Mass.
A. J. Manuel, Leeds University

QC21.2

L 578 — 1 : 3

1. Physics.

Theoretical Physics

An Advanced Text

Volume 3:

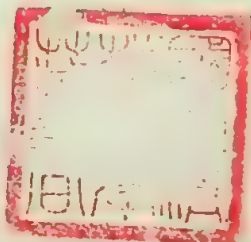
QUANTUM MECHANICS



Grigor'evich

Benjamin G. Levich, V. A. Myamlin and Yu. A. Vdovin

Institute of Electrochemistry
Academy of Sciences of the USSR, Moscow



1973

North-Holland Publishing Company – Amsterdam • London

Wiley Interscience Division
John Wiley & Sons, Inc. – New York

020497

© NORTH-HOLLAND PUBLISHING COMPANY, 1973

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior permission of the Copyright owner.

Library of Congress Catalog Card Number: 68 54501

ISBN North-Holland, complete set: 0 7204 0176 3

Vol. 3: 0 7204 0179 8

Printed in The Netherlands

Title of the Russian edition:

KURS TEORETICHESKOJ FIZIKI

Russian edition published by:

IZDATELSTVO 'NAUKA', GLAVNAJA REDAKCIJA,

FIZIKO-MATEMATIČESKOJ LITERATURY (MOSKVA, 1971)

Publishers:

NORTH-HOLLAND PUBLISHING COMPANY - AMSTERDAM

Sole Distributors for the Western Hemisphere:

WILEY INTERSCIENCE DIVISION

JOHN WILEY & SONS, INC. - NEW YORK

ISBN Wiley Interscience, Vol. 3: 0-471-53115-4

FOREWORD

The first Russian edition of 'Theoretical Physics', which appeared in 1962, has been widely used as a textbook.

Numerous comments from colleagues, lecturers and students have been taken into account in preparing this new edition, which is the first one in English and which will also appear as the second Russian edition.

The material has now been divided into 4 volumes covering the following subjects

Volume 1

- Part I Theory of the Electromagnetic Field
- Part II Theory of Relativity

Volume 2

- Part III Statistical Physics
- Part IV Electromagnetic Processes in Matter

Volume 3

- Part V Quantum Mechanics

Volume 4

- Part VI Quantum Statistics and Physical Kinetics

The rapid development of physics and the present wide interest in non-equilibrium and non-stationary processes has compelled us to expand the section on physical kinetics. It has also been transferred to the end of Volume 4 as it is practically impossible to expound this topic without using quantum mechanics.

Part IV – 'Electromagnetic Processes in Matter' – has been substantially revised. Interest in this field has increased recently, mainly in connection with the study of plasmas and plasma-like media, which now have sections devoted to them.

The methods of calculating electrostatic and direct-current fields, and other problems of classical electrodynamics in a medium, are covered very briefly as we have assumed that students will be able to consult the many monographs and handbooks on general physics, electrical- and radio-technology, and the equations of mathematical physics.

As for other modifications and additions, we should draw attention to the introduction of tensor notation, to new ideas in the theories of relativity and electromagnetic fields, the broadening of the introduction to the theory of probability, a brief presentation of the method of correlation functions in statistical physics, the exposition of the thermodynamic theory of ferromagnetism and the theory of propagation of electromagnetic waves in plasma. A number of paragraphs have been rewritten. We have tried to bring the content of the book even closer to the interests of present-day theoretical physics.

The general level of the book has been preserved and it is still intended to form an introduction to theoretical physics. Problems requiring the use of cumbersome or special mathematical apparatus are still excluded, and the most difficult sections are marked by an asterisk. These may be skipped at will, since there is no reference to them in the main text.

In conclusion we would like to express our gratitude to all those who helped us in preparing this book, in particular to A.M. Brodsky, A.M. Golovin, B.M. Grafov, R.R. Dogonadze, V.S. Krylov and especially V.S. Markin and V.V. Tolmachev. I.V. Savelyev discovered a number of misprints which have now been corrected.

L.D. Konkina helped us in editing the manuscript.

We are grateful to the readers and students who used the first Russian edition of the book for sending us their valuable comments which have been taken into account in this edition.

August 1970

FOREWORD TO VOLUME 3

This volume was written by Benjamin G. Levich in collaboration with Victor A. Myamlin and Yuri A. Vdovin. Dr. Anatol I. Naumov is the author of Chapter 15, *Fundamentals of the theory of elementary particles*. The authors express their deep gratitude to Dr. Naumov for this valuable assistance. The complete volume was written under the supervision of B.G. Levich.

October 1971

FOREWORD TO THE FIRST RUSSIAN EDITION

The continuous development of theoretical physics and the regular expansion of its areas of application create increasing demand for textbooks and manuals.

The rapid development and the complexity of the most recent experimental methods of physical investigation, and the corresponding development and extension of the mathematical apparatus of theoretical physics, have meant that one man usually cannot combine the two methods of investigation. The end of the 19th century and particularly the 20th century therefore saw physicists divided into 'experimentalists' and 'theoreticians', the latter studying physical laws by means of the mathematical methods of theoretical physics.

Obviously, a background in theoretical physics is essential in the education of experimental as well as theoretical physicists.

The experimental and theoretical methods of physical investigation have penetrated into a number of branches of science related to physics (physical chemistry, biophysics, geophysics, astrophysics, and so on) and into technology (metal physics and metallurgical science, thermophysics, electrical technology, radiotechnology, computation, the instrument-making industry etc.). Workers in these branches of science and technology also need a certain minimum knowledge of theoretical physics.

The compilation of a modern textbook on theoretical physics is inevitably associated with certain logical and methodological difficulties. It is impossible at present to divide theoretical physics into classical and quantum parts so that it is also impossible to divide it into separate chapters and sections. For example, the exposition of statistical physics without taking into account the quantum properties of atomic systems is impossible, for it would mean that the general theory remained without practical application. In the theory of electromagnetic processes in matter one has of necessity to make use of the ideas of statistical physics, and so on. It may be that the maximum consistency of composition would be obtained if the book were founded on

quantum mechanics but this is completely inadmissible in a book intended as an introductory treatise. Quantum mechanics requires a certain preparedness and the student must be convinced of the necessity of renouncing obvious classical representations. Compromise solutions, which have justified themselves during many years of teaching theoretical physics at the Moscow Engineering-Physical Institute and Moscow State University, are therefore inevitable.

The following general principles have been applied.

(1) The book is written as an introduction to theoretical physics so that aspects requiring the use of cumbersome or special mathematical apparatus have not been included.

(2) As it is to be used for a systematic study of the subject the course is a unique whole and all material necessary for understanding the later sections is contained in the earlier ones.

(3) It would not be feasible to elucidate experimental facts in addition to problems concerning purely theoretical physics. However, physics is a single science, and an attempt to expound the theoretical aspects without taking experiment into account would be quite wrong. The reader is assumed to have some basic experimental knowledge from university courses in general and atomic physics so that we have confined ourselves to references and, in a few instances, to a schematic description of basic experiments.

(4) The acquaintance assumed with general courses in general and atomic physics has allowed us to rely on a certain (very restricted) knowledge of quantum mechanics in our treatment of statistical physics.

(5) Classical mechanics usually forms a separate course so that this topic has been omitted although detailed reference has been made to handbooks of mechanics.

(6) The book similarly does not cover hydrodynamics, aerodynamics, the theory of heat transfer, or problems related to electrical- and radio-technology.

(7) Detailed reference is made to mathematical manuals. The mathematical apparatus utilized, except in the sections marked by an asterisk, is covered by the usual courses in analysis. In the case of quantum mechanics, however, the mathematical apparatus has been included, since it is of a specific character and is not taught in traditional mathematical courses.

(8) As the book is intended as a systematic course in theoretical physics no attempt has been made to achieve the same level of accessibility in all sections. It is a well-known fact that a student's comprehension and assimilation of difficult material increases as a course progresses, and that this is also true for the associated mathematical apparatus. Moreover, experi-

mental physicists will constantly encounter new problems in quantum mechanics which can only be handled using advanced methods of treatment. The section on quantum mechanics (Part V) therefore deals with some topics having a more advanced character than those in other sections. The analysis of applications of the kinetic equations is similarly treated rather extensively.

The uniqueness of the book's objectives has affected the content of individual sections, so that some topics in modern physics have been included at the expense of more traditional material.

Part I contains the foundations of the theory of the electromagnetic field in a vacuum, based on the system of Maxwell-Lorentz equations. A basic knowledge of electromagnetism is assumed. The focus of attention is the theory of radiation and the motion of charged particles in external fields.

In Part II, devoted to the theory of relativity, a four-dimensional form of representation is adopted which not only corresponds to the spirit of the theory but also predominates in contemporary literature. The problems of dynamics in the theory of relativity are treated in some detail. A number of the most recent applications of the theory of relativity, particularly those related to nuclear physics, are covered here for the first time in a textbook.

Part III is a revised version of Levich's 'Introduction to Statistical Physics' and treats statistical physics and the fundamentals of statistical thermodynamics. Classical thermodynamics would require too much space, and did not seem indispensable.

Part IV contains the theory of electromagnetic processes in matter. Relatively little attention is paid to problems in theoretical electrical- and radio-technology. The phenomenological theory of electric and magnetic properties of matter is analyzed in some detail, and the notion of the physics of the plasma state of matter is given.

In Part V the basic ideas of present-day relativistic quantum mechanics are included as well as the traditional problems of non-relativistic quantum mechanics. Applications to solid-state theory are considered at length.

Part VI contains the essential concepts of physical kinetics, which are not usually presented in a general course on theoretical physics.

The experience of teaching theoretical physics shows that the greatest difficulties are often encountered not in understanding new physical ideas but in the actual mathematical treatments. All mathematical operations have therefore been performed in sufficient detail.

For convenience we have presented a brief derivation of those formulae of

vector analysis which are encountered throughout, as well as the necessary data on Fourier integrals and δ -function theory.

The numbering of formulae and sections starts afresh in each Part and references to appendices have been given Roman numerals.

The author hopes that the readers, after making themselves familiar with the foundations of theoretical physics expounded in this book, will be able to proceed to a more profound study using the many-volume treatise of Landau and Lifshitz. The scientific and educational ideas of their work were of great influence on the author, who is a disciple of Landau.

Parts I–IV and Part VI were written by B.G. Levich. Part V was written by Y.A. Vdovin and V.A. Myamlin under the general scientific guidance of B.G. Levich. Chapter XV of Part V was written by A.I. Naumov.

The author expresses his gratitude to the colleagues who read the book and the manuscripts, and made a number of valuable remarks: B.M. Grafov, R.R. Dogonadze, V.A. Kiryanov, V.S. Krylov, V.S. Markin, V.P. Smilga, Y.A. Chizmadzhev and Y.I. Yalamov.

The creation of a textbook on theoretical physics sufficiently comprehensive in content and clear in presentation is a very complex task. The author is therefore conscious of the fact that shortcomings and errors will be discovered and would be grateful to receive an account of them which can be taken into consideration in the next edition of the book.

Theoretical Physics: Outline of Vols. 1—4

Volume 1

Part I Theory of the Electromagnetic Field

- Chapter 1 General theory of the electromagnetic field
- 2 The electrostatic field
 - 3 The quasistationary magnetic field
 - 4 The electromagnetic field of arbitrarily moving charges
 - 5 Radiation theory
 - 6 Electromagnetic field in a vacuum and electromagnetic wave scattering
 - 7 The motion of particles in electromagnetic fields

Part II Theory of Relativity

- Chapter 1 General principles of the theory of relativity
- 2 Relativistic mechanics
 - 3 Relativistic electrodynamics

Appendix I, II and III

Subject index

Volume 2

Part III Statistical Physics

- Chapter 1 The basic concepts of the theory of probability
- 2 The kinetic theory of gases

- 3 Statistical distribution
- 4 Statistical and phenomenological thermodynamics
- 5 Ideal gases
- 6 Systems of interacting particles
- 7 Crystals
- 8 The theory of fluctuations
- 9 Systems with a variable number of particles
- 10 Statistical distributions in quantum statistics and some of their applications

Part IV Electromagnetic Processes in Matter

- Chapter 1 Electromagnetic fields in matter
- 2 Electrostatics
 - 3 Direct electric current and the magnetic properties of matter
 - 4 Quasistationary electromagnetic fields
 - 5 High-frequency fields
 - 6 Matter in the plasma state

Appendix IV

Subject index

Volume 3 (for details see p. xv)

Part V Quantum Mechanics

- Chapter 1 The basic concepts of quantum mechanics
- 2 The Schrödinger equation
 - 3 The mathematical apparatus of quantum mechanics
 - 4 Motion in a centrally symmetric field
 - 5 The quasi-classical approximation
 - 6 The matrix form of quantum mechanics
 - 7 Perturbation theory
 - 8 Spin and identity of particles
 - 9 Applications of quantum mechanisms to atomic and nuclear systems
 - 10 The theory of diatomic molecules
 - 11 Scattering theory

- 12 The method of second quantization and radiation theory
- 13 Relativistic quantum mechanics
- 14 Some problems of quantum electrodynamics
- 15 Fundamentals of the theory of elementary particles

Subject index

Volume 4

Part VI Quantum Statistics and Physical Kinetics

- Chapter 1 Quantum statistics
 - 2 Physical kinetics
 - 3 The kinetic theory of gases and gas-like systems
 - 4 The time correlation function method and Onsager's theory
 - 5 Solid-state theory
 - 6 The kinetic properties of solids
 - 7 Interaction of radiation with a free-electron gas

Subject index



Contents of Volume 3

Part V Quantum mechanics

Chapter 1 The basic concepts of quantum mechanics

§ 1	The physical basis of quantum mechanics	3
2	The wave function	10
3	The principle of superposition. Expansion in plane waves	16
4	Uncertainty relations and the relationship between quantum mechanics and classical mechanics	19
5	The principle of causality in quantum mechanics	23

Chapter 2 The Schrödinger equation

6	The Schrödinger wave equation	26
7	The probability current density	30
8	A particle in a one-dimensional rectangular potential well	32
9	A particle in a three-dimensional rectangular potential well	36
10	The quantum-mechanical oscillator	37
11	The three-dimensional oscillator	43
12	Reflection from and penetration through a potential barrier	44
13	One-dimensional motion	52
14	The Schrödinger equation for a system of particles	55

Chapter 3 The mathematical apparatus of quantum mechanics

15	Linear operators	59
16	Eigenvalues and eigenfunctions of operators	63
17	Hermitian operators	65
18	The orthogonality and normalization of the eigenfunctions of Hermitian operators	67
19	Expansion in terms of eigenfunctions	69
20	Quantum-mechanical variables and operators	71

§21	The wave function and the probability of the results of measurements	73
22	Mean values	75
23	Commutation of operators	78
24	Heisenberg inequalities	82
25	Poisson brackets	84
26	Coordinate and momentum operators and eigenfunctions	86
27	The Hamiltonian operator	92
28	Stationary states	94
29	Integral form of the Schrödinger equation	96
30	The eigenvalues and eigenfunctions of the angular momentum operator and of the operator of the square of the angular momentum	101
31	Differentiation of operators with respect to time	108
32	Constants of the motion	111
33	Parity	114
34	The uncertainty relation for time and energy	117

Chapter 4 Motion in a centrally symmetric field

35	The Schrödinger equation	121
36	The free motion of a particle with given angular momentum	130
37	The spherical well	132
38	Motion in a Coulomb field	135

Chapter 5 The quasi-classical approximation

39	The limiting transition to classical mechanics	145
40	The solution of the Schrödinger equation near a turning point	150
41	Motion in a potential well in the quasi-classical approximation	155
42	Potential barrier penetration	158
43	Quasi-classical motion in a centrally symmetric field	162

Chapter 6 The matrix form of quantum mechanics

44	Operators and matrices	164
45	The fundamentals of matrix calculus	168
46	Geometric interpretation of the wave function. Canonical transformations	176

§ 47	The eigenfunctions and eigenvalues of an operator given in matrix form	180
48	Continuous matrices. The Dirac notation	183
49	The Schrödinger representation, the Heisenberg representation and the interaction representation	190
50	The linear harmonic oscillator	195
51	The matrix elements of the angular momentum operator	199
52	The addition of angular momenta	204

Chapter 7 Perturbation theory

53	The theory of time-independent perturbations	209
54	Perturbation theory in the presence of degeneracy	214
55	The theory of time-dependent perturbations	219
56	The transition of a system into new states under the action of perturbations	222
57	The adiabatic theory of perturbations	229
58	Perturbation theory in integral form	232

Chapter 8 Spin and identity of particles

59	The spin of elementary particles	234
60	Spin operators	236
61	The eigenfunctions of the operators of the components of the spin of a particle. The rotation matrix	240
62	The total angular momentum	248
63	The Pauli equation. The probability current density vector	250
64	The identity of particles. The principle of identity of particles. Symmetric and antisymmetric states	254
65	Wave functions for systems of fermions and bosons. The Pauli principle	257
66	The wave function of a system of two identical particles with spin $\frac{1}{2}$	261
67	Exchange interaction and the concept of the chemical and strong nuclear interactions	264

Chapter 9 Applications of quantum mechanics to atomic and nuclear systems

68	The helium atom	276
----	-----------------	-----

§ 69	The variational principle	280
70	The self-consistent field method (Hartree—Fock method)	282
71	The statistical model of the atom	285
72	The quantum numbers characterizing the states of electrons in atoms	293
73	The periodic system of the elements	299
74	The Zeeman effect	307
75	The Paschen—Back effect and the diamagnetism of atoms	312
76	Deuteron theory	314
77	Nuclear-shell theory	320

Chapter 10 The theory of diatomic molecules

78	The adiabatic approximation and the classification of electron terms	323
79	The hydrogen molecule. Ideas of the theory of chemical binding	326
80	The interaction of atoms at large distances	334
81	The comparison of molecular terms with atomic terms	337
82	Rotation and vibration of diatomic molecules	339

Chapter 11 Scattering theory

83	Scattering amplitude and cross section	345
84	The Born approximation	350
85	The scattering of fast charged particles by atoms	354
86	Partial wave scattering theory	357
87	Scattering by a spherical potential well (the concept of resonance scattering)	364
88	The elastic scattering of identical particles	368
89	The effect of polarization in scattering processes	371
90	The transition to the classical limit in the quantum scattering formulae	378
91	The general theory of inelastic scattering and the absorption of particles	385
92	The diffraction of fast neutrons by nuclei	391
93	The scattering of slow particles. The threshold approximation	394
94	The Breit—Wigner formula	396
95	The scattering matrix (<i>S</i> -matrix)	401
96	<i>S</i> -matrix and perturbation theory	408

§ 97	Analytic properties of the S -matrix	412
98	Time reversal and the principle of detailed balance	418

Chapter 12 The method of second quantization and radiation theory

99	Second quantization for systems of bosons and fermions	423
100	The quantum mechanics of the photon	431
101	The quantization of the radiation field	436
102	The interaction of an electron with radiation	439
103	The absorption and emission of light	442
104	Dipole transitions in atomic systems	446
105	Quadrupole and magnetic dipole radiation	448
106	Selection rules	449
107	The photoelectric effect	452
108	The scattering of light by atoms	456
109	The theory of the natural linewidth	462

Chapter 13 Relativistic quantum mechanics

110	The relativistic wave equation for a particle of zero spin	468
111	The charge density and probability current for particles of zero spin	470
112	The concept of the nuclear force field	473
113	The Dirac equation	476
114	The probability density and probability current in Dirac's theory	482
115	The solution of the Dirac equation for a free particle	483
116	The concept of the positron	487
117	The spin of particles described by the Dirac equation	492
118	The transition from the Dirac equation to the Pauli equation. The magnetic moment of a particle	494
119	The hydrogen atom in Dirac's theory	497
120	The invariance of the Dirac equation with respect to reflection, rotation and Lorentz transformation of coordinates	499
121	The laws of transformation of bilinear combinations made up of wave functions	502
122	The concept of weak interactions. Parity non-conservation	504
123	Two-component neutrino theory. The universal four-fermion interaction	509

Chapter 14 Some problems of quantum electrodynamics

§124	The Green's function of the Dirac equation	517
125	Green's function for a system of two particles	524
126	Feynman diagrams	527
127	The Compton effect	535
128	The shift of the terms of the hydrogen atom under the action of the vacuum field (the Lamb shift)	543

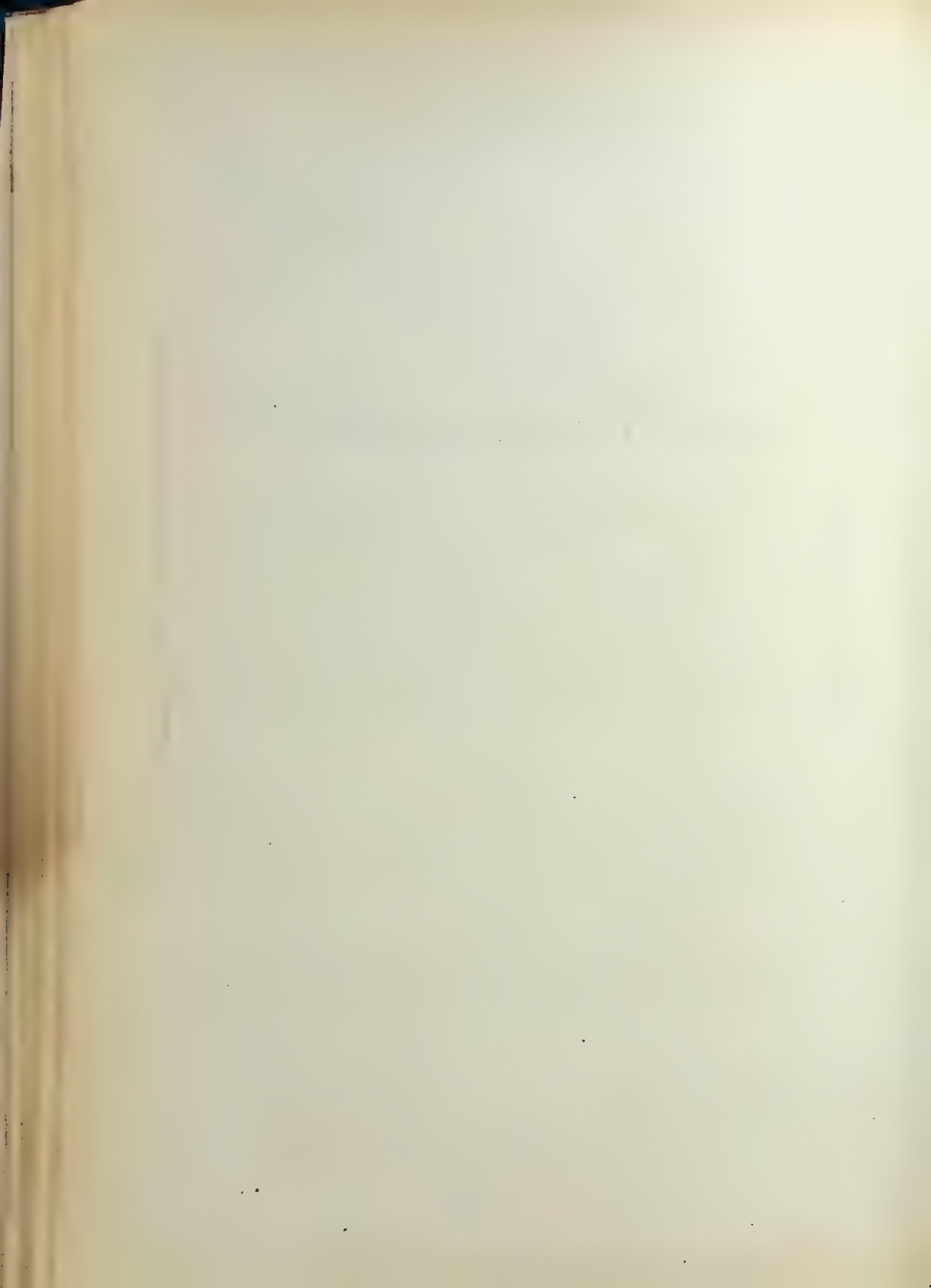
Chapter 15 Fundamentals of the theory of elementary particles

129	The classification and properties of elementary particles	549
130	The types of interactions of elementary particles	554
131	Symmetry groups in quantum mechanics	556
132	The isogroup $SU(2)$ and its representations	562
133	Isomultiplets of elementary particles	567
134	The wave functions of a system of nucleons and π -mesons	573
135	Isotopically invariant interaction	576
136	The scattering of nucleons and π -mesons	580
137	The unitary group $SU(3)$ and its representations	585
138	The eightfold way formalism and unitary multiplets	590
139	Some consequences of strict unitary symmetry	595
140	Some aspects of violated unitary symmetry	601
141	Composite models in the unitary symmetry scheme. Quarks	604
142	General appraisal of unitary symmetry	609

Subject index	613
---------------	-----

PART V

QUANTUM MECHANICS



The Basic Concepts of Quantum Mechanics

§1. The physical basis of quantum mechanics

Quantum mechanics, like all physical theories, arose in close connection with the development of a new field of experimental investigations. These investigations, which began with the study of black-body radiation at the turn of the century, were soon extended to the phenomena of the photoelectric effect, and subsequently to atomic systems. In this book we cannot give a chronological account of the history of the development of new concepts about the character of atomic processes which resulted in the creation of contemporary quantum mechanics. We can only point out that this was an agonizing search which required a great effort by some of the most outstanding physicists of our century. The creation of quantum mechanics has undoubtedly been the greatest triumph of contemporary science. The difficulties which the development of quantum mechanics encountered were associated with the fact that the properties of the particles constituting atomic systems are fundamentally different from the properties of macroscopic bodies. The laws of classical mechanics and classical electrodynamics turned out to be inadequate for the description of the behaviour of individual molecules and atoms, as well as elementary particles – electrons, protons, neutrons and so on. Henceforth we shall designate elementary particles, and sometimes

individual atoms and molecules, by the term microparticles. As we shall see in what follows, an outstanding feature of microparticles is the fact that their motion does not obey the laws of classical mechanics. To avoid confusion, we shall call particles whose motion does obey the laws of classical mechanics, corpuscles. We have earlier, in particular in the theory of the electromagnetic field and statistical physics, acquainted ourselves with a number of facts which indicate the inadequacy of classical concepts in the realm of atomic processes. Thus in statistical physics we have seen that the energy of individual atoms and molecules, i.e. the basic quantity characterizing their state, assumes discrete values.

A direct proof of the discrete character of the states of atomic systems was provided by the experiments of Franck and Hertz (1913). As is known, in these experiments a beam of electrons of a given energy entered a container filled with a gas. As a function of the accelerating potential, the electron current through the gas displayed a number of sharp minima. The position of these minima is determined by the properties of the atoms of the gas. This dependence of the current on the potential can be interpreted in the following way. In colliding with the atoms, the electrons transfer energy to the former when the energy of the electrons has a value equal to the difference between the energies of two possible states of the atom. The atom then makes transition to a state with a higher energy, an excited state, and a minimum appears in the electron current. If the energies of the electrons have other values, they undergo only elastic collisions. Thus the atom, as a whole, can only obtain definite amounts of energy from outside. This means that the internal energy of the atom has only discrete values or, in other words, the atom possesses a discrete energy spectrum. The discrete character of the energy states is related to the discrete character of atomic transitions. When the atom makes a transition from an excited state to a lower energy state a light quantum is emitted with an energy equal to the difference between the energies of the two states.

The energy of the atom is not the only quantity which can assume only discrete or, as it is said, quantum values. In the experiments of Stern and Gerlach it was shown that the angular momentum of an atom also possesses a discrete spectrum of values. In these experiments a beam of atoms was passed through a magnetic field \mathcal{H} which is nonuniform but constant in direction. Choosing this direction as the z -axis, one can write the expression $\mu_z(\partial\mathcal{H}_z/\partial z)$ for the force acting on the atom, where μ_z is the projection of the magnetic moment of the atom, onto the direction of the field. If it is assumed that the theorem of the proportionality between the magnetic moment and the angular momentum (see Part I, §22) is valid for atoms, then it follows that the mean

force is proportional to the value of L_z , where L_z is the projection of the angular momentum of the atom onto the direction of the field (see however ch. 8).

The experiment of Stern and Gerlach showed that the beam of atoms was deflected in the magnetic field, splitting into a number of separate beams. This means that the projection of the angular momentum of the atom onto the direction of the field can assume only discrete values. To every allowed value of L_z there corresponds a definite value of the force and a corresponding magnitude of the deflection in the nonuniform magnetic field. Thus each of the beams produced contains atoms with a given value of the quantity L_z .

The discrete character of the allowed values of the basic quantities which characterize the states of atomic systems profoundly contradicts all the concepts of classical mechanics. It follows from the general propositions of classical mechanics that an infinitesimal force causes an infinitesimal change in the state of a system. Hence all mechanical quantities depending on the state of that system, such as the energy, the momentum and so on, are continuous functions of the state. The discreteness of the states and the discontinuous changes of the states of microparticles directly contradicts this general principle.

The difficulty in understanding the properties of microparticles is increased by the fact that, in addition to the discreteness of certain quantities which characterize the state of particles, in a number of experiments the continuous nature of the same quantities was clearly shown. Thus, for example, the bremsstrahlung of electrons in the nuclear field has a continuous spectrum, which is indicative of a continuous change in the energy of the emitting particles.

It turned out that microparticles combine in a striking way the properties of ordinary particles (corpuscles) and the properties of waves. This fundamental property of microparticles is called the wave—particle duality.

The basic feature of corpuscles as studied in classical mechanics is that they have a definite spatial extent. The idealization of the corpuscle is a material point having no size and moving in a definite trajectory.

The properties of wave motion in classical physics are to a certain extent the opposite of the properties of corpuscular objects. A monochromatic wave, first of all, has an infinite extent in space. Hence it makes no sense to state that 'the monochromatic wave is located at a given point of space'. Also it is meaningless to speak of the trajectory of a monochromatic wave. The localization of a wave in space is inevitably associated with the production of a wave packet (see Part I, §35). The size of the wave packet is smaller, the larger the number of waves of different frequencies which form it. This prop-

erty of wave motion does not depend at all on their physical nature; it is valid for elastic, electromagnetic and other types of wave. Thus in classical physics localized corpuscles and waves delocalized in space are in a sense opposites.

It turns out that a combination of corpuscular and wave properties, inexplicable from the point of view of the ordinary ideas of classical physics, occurs in microparticles. More precisely, under certain conditions, microparticles behave as corpuscles, while under other conditions the same microparticles display purely wave properties. Finally, in certain experiments both corpuscular and wave properties manifest themselves simultaneously.

The wave-particle duality of the properties of microparticles was first discovered in experiments with light quanta. The wave properties of the electromagnetic field are sufficiently well established. We note only that Newton's corpuscular theory was able to compete successfully with the wave theory in explaining phenomena such as the rectilinear propagation and refraction of light. However, this theory was completely abandoned after the discovery of interference, diffraction and birefringence.

As to the corpuscular properties of the electromagnetic field, they are manifested in a particularly obvious way in the Compton effect (Part II, § 17; see also ch. 14). Indeed, this effect allows only the corpuscular interpretation. No considerations based on wave concepts can explain the appearance of recoil electrons: the incident electromagnetic wave cannot cause the motion of one of the atomic electrons without perturbing the motion of the remaining electrons. However, as we have seen in § 17 of Part II, the theory based on the concept of the collision of two particles, the incident photon and the atomic electron, describes the process correctly.

The corpuscular nature of light shows up in the same obvious way in the photoelectric effect, in the phenomenon of recoil when atoms are emitting radiation, etc. Thus the wave theory of light, which was successfully applied in considering a wide range of electromagnetic phenomena, turned out to be completely unsuitable for the explanation of a number of processes in which the corpuscular nature of light was manifested.

The situation was characterized briefly by stating that there is a dualism in the properties of the electromagnetic field. Sometimes light manifests its wave nature, and sometimes it behaves as a flux of photons.

The totality of experimental data showed that one must ascribe to every photon an energy E and a momentum p which are respectively equal to

$$E = \hbar\omega, \quad (1.1)$$

$$p = \frac{E}{c} = \frac{2\pi\hbar}{\lambda} = \frac{h}{\lambda}, \quad (1.2)$$

where h is the Planck constant h divided by 2π and equal to 1.054×10^{-27} erg·sec, and $\lambda = \lambda/2\pi$. Further, it turns out that the wave-particle duality appears not only in the case of photons, but for all microparticles.

The corpuscular properties of microparticles were discovered a relatively long time ago. They show up particularly clearly in observations with cloud chambers. As is known, microparticles in passing through a cloud chamber filled with a saturated vapour produce ionization along their path. The ions produced by the microparticles become the centres of condensation which can be observed directly in the form of tracks. Similarly, particles in moving through a thick layer of photographic emulsion leave a photographic image, i.e. a track. All this led to the idea that microparticles move in well defined trajectories and are similar in their properties to ordinary corpuscles. However, the experiments to be described below made it possible to establish that this was not so and that the wave-particle duality is a basic feature of all microparticles. But it should be stressed that the discovery of the wave properties of electrons, protons and other microparticles was preceded by the development of the concepts of quantum mechanics, in which the existence of the wave properties of microparticles was predicted theoretically.

Let us consider the following experiment. Individual electrons which have passed through a fixed accelerating field are let successively through a small opening in an impenetrable screen. After passing through the opening the electrons fall onto a photographic plate giving rise to blackening at the points of impact. If the electrons moved as corpuscles obeying the laws of classical mechanics and did not interact with the edge of the screen, then all of them would fall at the centre of the photographic plate, forming a black spot. In fact, the electrons must interact with the atoms of the screen. Since the latter are in thermal motion, this interaction has a random character. Hence it would be natural to expect the electrons to give rise to a blackening of the photographic plate similar to that caused by a beam of molecules coming out of a small opening. Namely, the number of electrons which are deflected from their rectilinear path and do not fall at the centre of the screen should depend on the magnitude of the deflection according to the normal law of errors. The intensity of the blackening, which is proportional to the number of electrons falling at a given point, should be expressed by the Gaussian distribution.

In fact, nothing of the kind is observed in experiment. If a large number of electrons are successively let through an opening, then the following is observed:

(1) There are zones on the photographic plate – ‘forbidden’ zones – in which electrons never arrive. These zones have the character of concentric rings of definite width;

(2) the zones in which electrons do arrive form a system of concentric rings alternating with the 'forbidden' rings.

By carrying out the experiment for a sufficiently long time, i.e. letting through a sufficient number of electrons, one can obtain blackened rings which are identical with those which arise when light is diffracted from a circular opening. Such a diffraction pattern is shown on the right of fig. V.1. In the drawing, white rings correspond to the blackened rings of the photographic plate. The curve of the intensity of electrons as a function of the angle of diffraction ϑ is shown on the left. The same result is also obtained in another arrangement of the experiment. Instead of letting electrons through one by one, a beam of electrons can be directed through the opening of the screen. The beam must be of a sufficiently low intensity, so that the interaction between electrons will be of no importance. When the beam of electrons passes through the opening of the screen, the distribution of the intensity of blackening of the photographic plate immediately appears in the form of a diffraction pattern*. Thus the motion of each individual electron differs fundamentally from that of a classical particle passing through the opening of the screen.

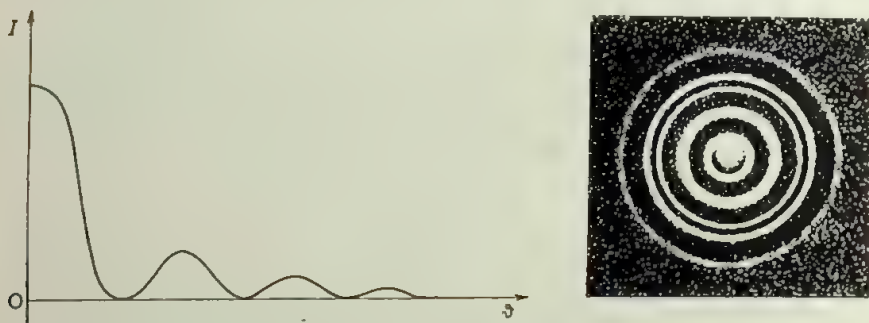


Fig. V.1

At first sight it may seem that the results of the observations described could be interpreted in the following way: for some unknown reason, not all possible trajectories of the motion of the electrons, but only certain allowed ones, can be realized in nature. The sum of these allowed trajectories deter-

* From the experimental point of view the second method is simpler. Such an experiment was carried out by Davisson and Germer in 1927, after the creation of the theory of quantum mechanics. The experiment with individual electrons was not carried out until 1948 by V.A. Fabrikant, L.M. Biberman and N. Sushkin.

mines the loci of incidence of electrons on the photographic plate. However, other experiments show this interpretation to be incorrect.

Let us consider an impenetrable screen with two openings. (The experiment discussed below is a schematization of a real experiment in which, instead of diffraction from a screen with two openings, the diffraction of electrons from a crystal lattice was observed.) If in turn each of the openings is covered, while individual electrons are successively let through the other, then after the passage of a large number of electrons the two diffraction patterns described above, with a central spot opposite to each of the openings, will arise on the photographic plate. We now uncover both openings and let electrons through. We assume that each of the electrons moves in a well defined allowed trajectory. Passing through one of the openings, the electron causes blackening at a definite point on the photographic plate. The final diffraction pattern produced by a large number of electrons should be a simple superposition of the intensities of the blackenings arising when electrons are let through one opening. In other words, we should obtain the same diffraction pattern as in the case of the alternative passage of electrons first through one opening and then through the other. In fact, however, the distribution of intensities of blackening is of a completely different character. The blackening of the photographic plate corresponds exactly to the pattern of diffraction from two openings. This means that there are no possible or allowed electron trajectories whatever. The electron, like a wave, possesses interference properties, and it would make no sense to try to establish through which one of the two uncovered openings a given electron 'in reality' passed.

We see that a certain wave motion is associated with the electron; the electron possesses wave properties. It is because of these wave properties that an individual electron passing through an opening can arrive at some regions of the photographic plate but not at others. In the passage through two openings the wave properties of an individual electron become apparent in the fact that its motion is affected by both openings. The allowed and forbidden regions of the photographic plate correspond to the dark and light zones of the diffraction pattern from two openings.

However, it would be incorrect, on the basis of the aforesaid, to try to identify the electron with a particular wave. If this were possible, then the darkening of the photographic plate on which the diffracted wave (the electron) falls would be a pale copy of the darkening produced by many electrons. An individual electron would immediately give the entire diffraction pattern.

We have stressed that experiment shows an individual electron to be inci-

dent at a definite point of the plate, like an ordinary corpuscle. The difference between an individual electron and a corpuscle lies in the fact that the loci of incidence on the photographic plate are determined by laws which are completely different from those determining the loci of incidence of a corpuscle. Thus, as is shown by the diffraction experiment, wave properties are inherent in every individual electron, but they manifest themselves in an obvious way only in the repetition of a large number of identical experiments (i.e. the successive passage of a large number of electrons).

We note that, although we have spoken above only of the electron, the same is also valid for other microparticles. Diffraction experiments have been carried out with neutrons, protons and other microparticles.

The quantum-mechanical treatment of the diffraction experiments described will be given in the next section. Here we stress once more that the wave-particle duality which had already been established for light quanta also becomes apparent in diffraction experiments with electrons.

Diffraction experiments make it possible to give an answer to the question: 'what is an electron — a wave or a corpuscle?' Here we use the terms 'wave' and 'corpuscle' in their usual classical meaning. The answer which follows directly from the experiments described is that the electron is neither a wave (otherwise a single electron would give the entire diffraction pattern), nor a corpuscle moving in a definite trajectory (which contradicts the experiment with two openings). The electron is a microparticle possessing specific properties.

§2. The wave function

The fact that the electron possesses wave properties shows that the electron is to be compared with a certain wave field. We shall call the amplitude of this wave field, which depends on the spatial coordinates and time, the wave function $\psi(x, y, z, t)$. For brevity it is sometimes also called the ψ -function.

The physical interpretation of the wave function (which was first given by M. Born) is the following: the quantity $|\psi(x, y, z, t)|^2 dV$ is proportional to the probability that the electron will be found at the instant of time t in a volume element dV in the neighbourhood of the point x, y, z .

Denoting this probability by dW , we have

$$dW \sim |\psi(x, y, z, t)|^2 dV. \quad (2.1)$$

This interpretation is based on the following reasoning. In the experiments

with the passage of individual electrons through one or two openings we have seen that the locus of incidence of the electron on the photographic plate is to a certain degree random. The electrons will fall completely randomly at one or other point of the diffraction ring to be formed. Hence the behaviour of an electron must be characterized by a certain probability function. The intensity of blackening of the photographic plate at a given spot is proportional to the number of electrons which fall on this spot. On the other hand, it is clear that this probability function must be connected with the properties of the wave field. Only in this case can the probabilistic character of the blackening of the photographic plate at a given spot be matched with the strict spatial distribution of the bands of blackening. That is, the random character of the incidence of the electron at a given point can be matched with its wave properties only by assuming that the probability of finding the electron at the given point is proportional to the intensity of the wave field $|\psi|^2$. This relation is just that given by formula (2.1).

The physical interpretation of the wave function given by formula (2.1) clearly shows that the wave field $\psi(x, y, z, t)$ is fundamentally different from other wave fields known in classical physics. This is particularly apparent from the fact that only the quantity $|\psi|^2$ has a direct physical meaning. The wave function itself, in general, can be a complex quantity. Furthermore, the wave functions ψ and $A\psi$, where A is an arbitrary constant, correspond to one and the same physical state of the particle, since by virtue of the definition (2.1) the two wave functions lead to one and the same space-time distribution of the probability of finding the particle.

By virtue of the theorem of addition of probabilities (see § 2 of Part III) the definition (2.1) can be supplemented by the following normalization condition:

$$\int |\psi(x, y, z, t)|^2 dV = 1, \quad (2.2)$$

where the integral on the left, taken over all space, is the probability of finding the particle at time t at any point of space. This probability is naturally equal to unity. Wave functions ψ satisfying the normalization condition are said to be normalized. For normalized wave functions the relation (2.1) can be rewritten in the form

$$dW = |\psi(x, y, z, t)|^2 dV = \rho(x, y, z, t) dV, \quad (2.3)$$

where $\rho(x, y, z, t)$ is the probability density. The probability $W(V, t)$ of finding the particle in a given finite volume V at an instant of time t , is, according

to the theorem of addition of probabilities

$$W(V, t) = \int_V dW = \int_V |\psi(x, y, z, t)|^2 dV. \quad (2.4)$$

The condition (2.2) cannot be satisfied in the case where the integral $\int |\psi|^2 dV$ is divergent. This can occur, in particular, if the square of the modulus of the wave function, $|\psi|^2$, does not tend to zero at infinity. Physically this means that there is a finite probability of finding the particle at every point of space. In §18 it will be shown how the normalization of the wave function is to be carried out in this case.

We note that the wave function normalized by the condition (2.2) is defined with an accuracy to within the factor $e^{i\alpha}$, where α is any real number, in view of the equality $|e^{i\alpha}|^2 = 1$.

In addition to the wave function of a single microparticle it is also necessary to introduce the idea of the wave function of a system of microparticles. Let there be a system of N particles interacting with each other according to an arbitrary law. This system of particles can be represented by the wave function

$$\psi(x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_i, y_i, z_i, \dots, x_N, y_N, z_N, t),$$

where i is the index of the particle.

In the further construction of quantum mechanics we shall proceed from the assumption that there is no difference of principle between the description of an individual microparticle and a system of microparticles and that the interpretation of $\psi(x, y, z, t)$ and $\psi(x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_N, y_N, z_N, t)$ must be one and the same. In other words, the physical meaning of the wave function of a system of N particles lies in the fact that the quantity

$$dW \sim |\psi(r_1, r_2, \dots, r_N, t)|^2 dV_1 dV_2 \dots dV_N \quad (2.5)$$

gives the probability that at a certain instant of time t the first particle be found in the volume element dV_1 surrounding the point r_1 , the second particle in the volume element dV_2 surrounding the point r_2 , and so on. For brevity, r_i here denotes the totality of coordinates (x_i, y_i, z_i) . We note that on the basis of the theorem of addition of probabilities the quantity

$$dW_1 \sim dV_1 \int \psi(r_1, r_2, \dots, r_N, t) dV_2 \dots dV_N \quad (2.6)$$

represents the probability of finding the first particle in the volume element dV_1 for any distribution of the remaining particles of the system (the integration being carried out with respect to the coordinates of the latter particles). It is obvious that the probability dW_1 given by formula (2.6) must be identi-

cal with the definition (2.3). Analogous relations can also be written for other particles of the system. Thus dW gives the probability of finding a given configuration of the system in space.

The normalization condition for the wave function of a system of N particles has the form

$$\int |\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, t)|^2 dV_1 dV_2 \dots dV_N = 1. \quad (2.7)$$

It is clear that the wave function of a system of N particles is not normalized in the real three-dimensional space but in the $3N$ -dimensional configuration space.

In view of the similarity of principle between the wave function of one particle and that of a system of particles, we shall always denote the wave function by the symbol ψ . For the sake of brevity the whole set of coordinates is sometimes denoted by x .

It follows from the aforesaid that the quantity $|\psi|^2$ must be interpreted as a probability not in real space but in configuration space. At the same time, the introduction of the wave function of a system of particles confirms in a particularly obvious way the impossibility of interpreting the wave function as a quantity which describes a wave motion similar to an electromagnetic or an acoustic wave propagating in real space. Indeed, every wave motion in real space is characterized by the set of three variable coordinates and time. However, the wave function of a system of N particles depends on $3N$ coordinates and time. Hence in interpreting the ψ -function as an ordinary wave one would have either to renounce the assumption of the unique meaning of the wave function of one microparticle and of a system of microparticles, or to introduce the hypothesis of the existence of a real multi-dimensional space. Both are in flagrant contradiction with all experimental data.

Let us consider the important particular case of a system of noninteracting particles. Then finding the i th particle in the volume element dV_i , the k th particle in the volume element dV_k and so on, must be independent events. On the basis of the theorem of multiplication of probabilities, formula (2.5) can be written in this case as follows:

$$\begin{aligned} dW &= dW_1 dW_2 \dots dW_N = \\ &= |\psi_1(\mathbf{r}_1, t)|^2 dV_1 |\psi_2(\mathbf{r}_2, t)|^2 dV_2 \dots |\psi_N(\mathbf{r}_N, t)|^2 dV_N. \end{aligned}$$

This means that the wave function of a system of noninteracting particles is equal to

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, t) = \psi_1(\mathbf{r}_1, t) \psi_2(\mathbf{r}_2, t) \dots \psi_N(\mathbf{r}_N, t). \quad (2.8)$$

The case where the system consists of identical microparticles (for example, of electrons or protons, and so on) will, in particular, be considered later (see §64).

Before trying to construct the wave function for the simple case of the motion of a microparticle, it is necessary to make the following very important remark. At first sight it might be assumed that it is necessary to introduce into physics new notions in order to describe the states of microparticles, which represent completely new objects as regards their physical nature. It turns out, however, that this is not so. The state and the nature of the motion of microparticles can to a certain degree be characterized by the quantities and terms of classical physics. This is pointed out by the wave-particle duality of microparticles, which consists in the fact that in certain experiments microparticles manifest themselves as objects with a wave nature, while in other experiments they behave as ordinary corpuscles.

When we introduced the statistical interpretation of the wave function we already assumed, to a certain degree, that the concepts of classical mechanics are applicable to microparticles. Indeed, the statement that 'a microparticle can be found in a volume element dV ' already implies the assumption that the classical approach is possible for the description of its state by defining its position in space. If the microparticle were in all respects similar to a wave, then the statement of the problem, 'where can the microparticle be found', would make no sense. On the other hand, the presence of the diffraction pattern makes it possible, under certain conditions, to associate the microparticle with another classical notion, a definite wavelength λ , and to speak of the wavelength corresponding to the wave function of the particle.

Classical ideas, such as the position of the particle, the wavelength and so on, can only be applied to microparticles within certain limits. We shall dwell in detail on this in §4. The most important thing is not the fact that in describing microparticles the notions of classical physics are of limited applicability, but the fact that they can and must be used in describing new objects which are so unlike ordinary macroscopic bodies or waves.

We shall assume that the state of an electron moving freely in space can be characterized by an energy E and a momentum \mathbf{p} . Then the relation between the energy and momentum is given by the classical formula

$$E = \frac{|\mathbf{p}|^2}{2m}. \quad (2.9)$$

We assume that a beam of electrons which has passed through a strictly defined accelerating potential difference and has acquired a definite energy enters a diffraction arrangement (in practice this arrangement is usually a

crystal lattice). Formula (2.9) allows us to speak of a definite momentum of the electron.

On the other hand, knowing the diffraction pattern one can (see §36 of Part IV) find the wavelength λ corresponding to the electron. It turns out that the following relation exists between the quantities λ and p :

$$p = 2\pi\hbar/\lambda = \hbar k. \quad (2.10)$$

The relation (2.10), which was first proposed in 1924 by de Broglie on the basis of theoretical considerations, is called de Broglie's formula. The wave associated with the motion of a microparticle is called the de Broglie wave.

We see that de Broglie's formula is the same as formula (1.2) for light quanta. The frequencies corresponding to the de Broglie waves cannot be determined directly in an experimental way. However, it is natural to assume that the relation between the energy and frequency which holds for light quanta is also applicable to the de Broglie waves, i.e.

$$E = \hbar\omega. \quad (2.11)$$

Based on the relations (2.10) and (2.11), we write the wave function of a free particle in the form of a plane monochromatic wave

$$\psi_{\mathbf{p}}(\mathbf{r}, t) = Ae^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} = Ae^{(i/\hbar)(\mathbf{p} \cdot \mathbf{r} - Et)} \quad (2.12)$$

Later it will be explained why $\psi_{\mathbf{p}}(\mathbf{r}, t)$ should be written in the form of an exponential function, and not in the form of a sine or cosine (see §6). The constant A is determined by the normalization condition (see §26).

By definition \mathbf{k} is the wave vector

$$\mathbf{k} = \frac{1}{\hbar} \mathbf{p}. \quad (2.13)$$

By means of formulae (2.9), (2.10) and (2.11) one can find the dispersion law of de Broglie waves

$$\omega = \frac{1}{\hbar} E = \frac{p^2}{2m\hbar} = \frac{\hbar k^2}{2m}. \quad (2.14)$$

The corresponding phase velocity and group velocity are equal to

$$v_{\text{ph}} = \frac{\omega}{k} = \frac{\hbar k}{2m}, \quad (2.15)$$

$$v_{\text{gr}} = \frac{d\omega}{dk} = \frac{\hbar k}{m} = \frac{1}{m} p. \quad (2.16)$$

Formula (2.16) shows that the group velocity of de Broglie waves is the same as the ordinary velocity of macroscopic particles. If we took the quantity v_{gr} as the initial expression for the velocity of the particle, then the relation between the energy and frequency (2.11) could be obtained as a consequence of this definition. The phase velocity v_{ph} of de Broglie waves has no direct physical meaning. This becomes particularly clear if one makes use of the relativistic expression for the relation of the energy of the momentum of the particle

$$E = (p^2 c^2 + m^2 c^4)^{\frac{1}{2}} = (\hbar^2 c^2 k^2 + m^2 c^4)^{\frac{1}{2}} = \hbar \omega .$$

Then

$$v_{ph} = \frac{\omega}{k} = \frac{E}{\hbar k} = \left(c^2 + \frac{m^2 c^4}{\hbar^2 k^2} \right)^{\frac{1}{2}} > c ,$$

i.e. v_{ph} is larger than the velocity of light.

§3. The principle of superposition. Expansion in plane waves

We have considered above the phenomenon of the diffraction of electrons incident on a screen with a limited number of apertures. In fact, however, what is observed is the diffraction of electrons from a crystal lattice. This case is not only of practical interest, but is also very important from the theoretical standpoint.

We have seen in §36 of Part IV that the selective reflection of X-rays takes place when the Bragg–Woolf condition is fulfilled. If we dispense with unessential details, it turns out that the diffraction of electrons from a crystal lattice is analogous to the diffraction of X-rays. As a result of the scattering of electrons from a single-crystal a number of selective reflections occurs. Each selective reflection corresponds to a definite momentum or, by virtue of (2.10), to a definite wavelength. Thus the crystal lattice is a device which resolves the initial polychromatic beam of electrons into a number of beams each of which corresponds to electrons of a definite wavelength.

We have already pointed out that the experiment with an electron beam is equivalent to the whole set of successive measurements with a large number of electrons which are in identical external conditions. Hence the diffraction grating plays the role of a device which analyzes the initial state of the micro-particle, resolving it into a set of individual states with definite values of the momentum. Since to each state with a definite momentum there corresponds a plane wave of the form (2.12), then consequently, the wave function

describing the initial state of the electron incident on the grating can in general be written in the form of a superposition of plane waves, i.e.

$$\psi(x, y, z, t) = \int_{-\infty}^{\infty} c(p_x, p_y, p_z) \psi_{\mathbf{p}}(x, y, z, t) dp_x dp_y dp_z. \quad (3.1)$$

Physically this means that the wave function of an electron in an arbitrary state can be considered as a superposition of the wave functions corresponding to states with a definite value of the momentum.

Hence it is not surprising that one electron (or other microparticle) can be in a definite state without having a definite value of the momentum. Although the concept of momentum, applied to the microparticle from classical mechanics, can be used in quantum mechanics, the state of a microparticle is not defined by the same laws as the state of a particle in classical mechanics. In later sections we shall come back to the discussion of the problem of characterizing the states of microparticles.

Choosing the coefficient A of formula (2.12) in the form (see §26)

$$A = \frac{1}{(2\pi\hbar)^{\frac{3}{2}}}, \quad (3.2)$$

we have

$$\psi(\mathbf{r}, t) = \int_{-\infty}^{\infty} c(\mathbf{p}) \frac{1}{(2\pi\hbar)^{\frac{3}{2}}} e^{(i/\hbar)(\mathbf{p} \cdot \mathbf{r} - Et)} d\mathbf{p}. \quad (3.3)$$

From the mathematical point of view, formula (3.3) represents the expansion of the function $\psi(\mathbf{r}, t)$ in a Fourier integral. The amplitude $c(\mathbf{p})$ shows the weight with which the state $\psi_{\mathbf{p}}$ is involved in the state described by the wave function $\psi(\mathbf{r}, t)$. According to the Parseval equality (see, in Volume 1, Appendix II eq. II.9), we have

$$\int |\psi|^2 dV = \int |c|^2 d\mathbf{p}. \quad (3.4)$$

When choosing the coefficient A in the form of (3.2) the equality (3.4) contains no numerical coefficients. It is natural to assume that $|c(\mathbf{p})|^2$ can be related to the density $\rho(\mathbf{p})$ of the probability that the value of the momentum of the particle in the state $\psi(\mathbf{r}, t)$ be equal to \mathbf{p} . Namely, it is natural to assume that

$$\rho(\mathbf{p}) = |c(\mathbf{p})|^2. \quad (3.5)$$

We write the equality sign and not the proportionality sign \sim , because the

function ψ is normalized by the condition (2.2). In this case the following equality holds (see (3.4)):

$$\int |c(p)|^2 dp = 1. \quad (3.6)$$

We shall return to the discussion of formula (3.5) in §21.

Equation (3.3) is a particular case of one of the most important propositions of quantum mechanics – the principle of superposition. This principle amounts to the following: if the quantum system can be in states described by functions $\psi_1, \psi_2, \dots, \psi_n$, then the linear combination (superposition) of the wave functions ψ_n

$$\psi = \sum_n c_n \psi_n, \quad (3.7)$$

where c_n are arbitrary constants, is also a wave function describing one of the possible states of the system. The importance of the principle of superposition, in particular, lies in the fact that it restricts the possible equations for the determination of ψ to linear equations (see §6). If the index n characterizing the state runs over a continuous sequence of values, then the summation in formula (3.7) must be replaced by integration. Later we shall come back to the discussion of the notion ‘the state of a quantum system’ and of the meaning of the coefficients c_n (see §21 and §23).

As an example of the application of the principle of superposition, let us consider a free particle whose momentum has no strictly defined value but can lie in a small interval Δp about the value p_0 , namely $p_0 - \Delta p \leq p \leq p_0 + \Delta p$. For simplicity we consider the case of uniform motion. According to (3.3) the wave function of the electron can be written in the form

$$\begin{aligned} \psi(x, t) &= \int_{p_0 - \Delta p}^{p_0 + \Delta p} c(p) \frac{1}{(2\pi\hbar)^{\frac{1}{2}}} e^{i(i/\hbar)(px - Et)} dp = \\ &= \left(\frac{\hbar}{2\pi}\right)^{\frac{1}{2}} \int_{k_0 - \Delta k}^{k_0 + \Delta k} c'(k) e^{i(kx - \omega t)} dk. \end{aligned} \quad (3.8)$$

For uniform motion the coefficient A is equal to $(2\pi\hbar)^{-\frac{1}{2}}$ (see §26). In accordance with the results of §35 of Part I, the expansion of (3.8) is expressed by a formula which is the same as (35.1) of Part I:

$$\psi(x, t) = \left(\frac{\hbar}{2\pi}\right)^{\frac{1}{2}} 2c'(k_0) \frac{\sin \left[\Delta k \left(x - \left(\frac{d\omega}{dk} \right)_0 t \right) \right]}{\left[x - \left(\frac{d\omega}{dk} \right)_0 t \right]} e^{i(k_0 x - \omega_0 t)}. \quad (3.9)$$

Formula (3.9) shows that the superposition of wave functions which correspond to nearby values of the momenta (or the wave numbers) leads to the formation of a wave packet propagating with the group velocity

$$v_{gr} = \left(\frac{d\omega}{dk} \right)_0 = \frac{1}{m} p_0.$$

It is clear from the form of the wave function (3.9) that the probability of finding the microparticle at a point x at an instant of time t , which is proportional to $|\psi(x, t)|^2$, has a sharp maximum moving with the velocity v_{gr} .

It should be stressed that eq. (3.9) is of an approximate character. Taking account of the subsequent terms of the expansion of the function $\omega(k)$ would lead to an expression for the wave packet whose width would increase with time. Such a wave packet is said to be spreading. The spread of the packet follows directly from the fact that each wave forming the packet moves with its phase velocity $v_{ph} = \omega/k = \hbar k/2m$.

When quantum mechanics first appeared attempts were made to identify the electron with a wave packet made up of de Broglie waves. However, the spread of the wave packet is indicative of the unsoundness of such a treatment. Furthermore, if the electron represented a wave packet, then, as in the case of a single wave, it would be impossible to account for the experiment on the diffraction of individual electrons.

§4. Uncertainty relations and the relationship between quantum mechanics and classical mechanics

We shall use the representation of the wave function in the form of a wave packet for the discussion of a fundamental problem. The question is to what extent and with what degree of accuracy use can be made of the concepts of classical mechanics in their application to microparticles. Here we restrict ourselves to the consideration of the concepts of the momentum and position of a particle in space. In §24 this problem will be studied in full.

We have seen in §35 of Part I that the wave packet possesses a spatial extent given by formula (35.7) of Part I. Applying this to the wave packet (3.9), in which we are interested here, this formula can be written in the form

$$\Delta p_x \Delta x \sim \hbar.$$

Since a spread of the packet takes place, which was not taken into account

in deriving this formula, it can be written more correctly as follows:

$$\Delta p_x \Delta x \gtrsim \hbar. \quad (4.1)$$

The numerical factor in formula (4.1) will be defined more precisely in §24. In a similar way one can write the relations for the remaining two coordinate and momentum components:

$$\Delta p_y \Delta y \gtrsim \hbar, \quad (4.2)$$

$$\Delta p_z \Delta z \gtrsim \hbar. \quad (4.3)$$

Formulae (4.1)–(4.3) are called Heisenberg's uncertainty relations. We discuss the meaning of these inequalities by proceeding from the probabilistic interpretation of the wave function. If the width of the wave packet is equal to Δx , then according to what was said in the preceding section the measurements of the coordinates of the electron will show that with a very high probability it will be found in the region of space Δx . In this sense it can be said that the coordinate of the electron is determined to within an accuracy of Δx . However, the electron found in the region Δx is not described by a plane wave and has no definite value of the momentum. To form a wave packet of width Δx it was necessary to form a superposition of plane waves with momenta in the interval $p_0 - \Delta p_x \leq p_x \leq p_0 + \Delta p_x$, where Δp_x is determined by formula (4.1). This means that the measurements of the momentum of an electron localized in the region Δx will lead to values of the momentum which lie in the interval mentioned. In other words, the uncertainty Δx in the value of the coordinate of an electron localized in the region Δx and the uncertainty Δp_x in the value of its momentum are connected by the relation (4.1). The smaller the width of the packet Δx , the larger Δp_x . Conversely, if the momentum interval Δp_x is defined, then formula (4.1) shows that the particle will be found with a very high probability in a region of space $\Delta x \geq \hbar/\Delta p_x$.

It follows from the inequality (4.1) that the values of Δx and Δp_x cannot simultaneously be equal to zero. This means that the x -coordinate and the momentum p_x associated with it cannot simultaneously have sharp values. Thus the classical concepts of the spatial position and the value of the momentum are applicable to microparticles only within definite limits given by Heisenberg's relations. Any attempt to apply simultaneously the concepts of the momentum and the coordinate to a microparticle with an accuracy higher than that given by the uncertainty relations makes no sense. This fact is associated with the very nature of microparticles, with their wave–corpuscle properties.

In this connection the reader should be warned against the erroneous assumption of certain authors that Heisenberg's uncertainty relations give that degree of accuracy with which the coordinates and momentum of microparticles can be determined within the framework of quantum mechanics. In their opinion, for a more accurate simultaneous determination of the coordinates and momenta a further development of theory is necessary.

In reality this is not so. Microparticles are completely new objects, by no means classical, with their characteristic properties and laws of motion. As we have already pointed out, a distinctive feature of microparticles is the dualism of the corpuscular and wave properties which they manifest. It follows from diffraction experiments that particles have no definite trajectory. Hence it is impossible to describe the motion of a particle by giving an accurate value of the coordinate and momentum at every instant of time, as is done in classical mechanics. However, one can indicate, with a certain degree of accuracy, the magnitude of that region of space in which the particle will be found with a very high probability, and the interval of those values of the momentum which it possesses at that time. The value of these quantities is given by Heisenberg's uncertainty relations.

It should be noted that when the particle has a definite value of the momentum, $\Delta p_x = 0$, then according to (4.1) its position is completely indefinite, i.e. $\Delta x \rightarrow \infty$. Indeed, a state with definite momentum is described by a plane de Broglie wave. For such a wave the square of the modulus $|\psi_p|^2$ is constant, i.e. the particle can be found with the same probability at any point of space.

On the other hand, if a definite position of the particle at a given instant of time is given, then its momentum is completely indefinite. It may seem that the relation obtained is in contradiction with the existence of distinct tracks of particles in a cloud chamber or on a photographic plate. However, this contradiction is only apparent. Indeed, the track of an electron in a cloud chamber represents liquid drops formed on the ions produced by the electron. The size of the drops gives the degree of accuracy with which the coordinate of the particle can be fixed. Since the size of the drops is of the order of 10^{-4} cm, the uncertainty in the coordinate of the electron is also of the order of 10^{-4} cm. Consequently, the uncertainty in the corresponding momentum component $\Delta p_x \sim \hbar/\Delta x \sim 10^{-23}$ g · cm · sec⁻¹. Since the mass of the electron is equal to $\sim 10^{-27}$ g, then the uncertainty in the velocity component perpendicular to the track will be equal to $\Delta v_x = m^{-1} \Delta p_x = 10^4$ cm · sec⁻¹.

But tracks in a cloud chamber are produced only by fast electrons having a velocity v of the order of $\geq 10^9$ cm/sec. Hence we see that under these condi-

tions $\Delta v_x \ll v$ and one can speak approximately of the motion of the particle along a trajectory in the cloud chamber.

Heisenberg's uncertainty relation written in the form

$$\Delta v_x \Delta x \gtrsim \frac{\hbar}{m} \quad (4.4)$$

shows that the concepts of classical physics turn out to be applicable with a degree of accuracy which is higher, the larger the mass of the particle. In view of the smallness of the quantum constant \hbar , the uncertainty in the values of the coordinate and velocity become negligibly small for particles of a macroscopically small but still not atomic size.

Let us, for example, have a body of the size of about 1 micron and with a mass of only 10^{-10} g. Then (4.4) gives $\Delta v_x \Delta x \sim 10^{-17} \text{ cm}^2 \cdot \text{sec}^{-1}$. If, for example, the position of the body is determined with an accuracy of 10^{-6} cm (1/100 of its size), then $\Delta v_x \sim 10^{-11} \text{ cm} \cdot \text{sec}^{-1}$. The velocity of the Brownian motion of a particle of a mass of 10^{-10} g amounts to $\sim 10^{-4} \text{ cm} \cdot \text{sec}^{-1}$. We see that the error in the velocity, which is associated with the uncertainty relation, is already negligibly small for such a small body. Even more so, is it of no importance for macroscopic bodies.

The estimates given illustrate a general important proposition of quantum mechanics which is called the correspondence principle: in passing to the limit $\hbar \rightarrow 0$, i.e. in assuming that the effects proportional to the quantum constant can be disregarded, the laws and relations of quantum mechanics go over into the corresponding laws and relations of classical mechanics. (For more detail on the transition to classical mechanics see ch. 5.) In particular, for particles of large mass the ratio \hbar/m is so small that in practice the coordinate and velocity have definite values. Such a particle has a trajectory along which it moves in accordance with the laws of classical mechanics. The importance of the correspondence principle lies in the fact that it serves as a method of finding the quantum-mechanical analogues of classical quantities. Quantum mechanics implies classical mechanics as a certain limiting case corresponding to $\hbar \rightarrow 0$ (for other conditions of this transition see ch. 5). From the correspondence principle it is possible to establish the relation between certain quantum-mechanical quantities and the concepts of classical mechanics.

Besides the reasoning presented above, the uncertainty relations are often obtained from a discussion of the possible degree of accuracy of the determination of the coordinate and momentum of a microparticle in different experiments which are in principle feasible. We shall not dwell on the analysis of these examples, because a strict derivation of the uncertainty relations will be given in §24.

It should be noted that if the region of possible motion of a microparticle is given, for example the size l of the atom or the nucleus, then the uncertainty relations make it possible to qualitatively estimate the values of its momentum and energy. Indeed, the absolute value of the momentum is of the same order of magnitude as its uncertainty $\Delta p \sim \hbar/l$. Consequently, $p \gtrsim \hbar/l$, and the energy of the particle is

$$E = \frac{p^2}{2m} \gtrsim \frac{\hbar^2}{2ml^2}. \quad (4.5)$$

We see that the energy increases with decreasing region of localization. For example, for an electron in the atom l is of the order of the size of the atom, i.e. of the order of 10^{-8} cm. Substituting this value into (4.5), we find the energy of the electron in the atom to be $E \gtrsim 10$ eV. This is of the correct order of magnitude.

Further, let us consider a proton or a neutron in a nucleus. The size of the nucleus is of the order of 10^{-12} cm. Setting $l \sim 10^{-12}$ cm and taking into account that the mass of the nucleon is $m \sim 10^{-24}$ g, we estimate the energy E to be $E \gtrsim 1$ MeV. This estimate is also in agreement with experimental data.

§5. The principle of causality in quantum mechanics

We have seen in the preceding section that the concepts of classical physics are applicable to microparticles only within certain limits. The question naturally arises: why can and must we, as a matter of fact, describe the motion of microparticles in terms of classical physics? The necessity of introducing classical concepts into quantum mechanics is associated with the following important fact: the explanation of the properties and laws of motion of micro-objects is possible only by setting them in interaction with macroscopic bodies. A macroscopic body interacting with microparticles is called an apparatus. The process of interaction between the apparatus and a microparticle is called measurement.

Of course, the apparatus in this sense of the word is not necessarily an artificial device for registering the properties of microparticles. The apparatus is any body which can change its state as a result of an interaction with micro-objects and which is described with a sufficient degree of accuracy by the laws of classical physics. The process of interaction of the apparatus with a microparticle (a measurement) is an objective process taking place in space and time. However, since any scientific information can only be based on the fact and character of the interaction mentioned, all characteristics of micro-

particles must be directly connected with the properties of their interaction with macroscopic bodies. This just means that the description of microparticles must necessarily imply, if only partially, the concepts of classical physics. Of course, there may also exist characteristics and properties of microparticles which manifest themselves in interactions with apparatus but have no classical analogue. We shall see, for example in ch. 8, that such a characteristic is the spin of microparticles.

The interaction between microparticles and macroscopic bodies differs, essentially of course, from the interaction of macroscopic bodies with each other. Namely, for the interaction between two macroscopic bodies, one of which is playing the role of the apparatus, one can always assume that the reaction of the apparatus on the body is as small as one wishes or, if one likes, one can take it accurately into account. Hence it is said that the effect of the apparatus does not change the state of the macroscopic object.

The situation is different in the case of the interaction between physical objects of different natures, i.e. a microparticle and a macroscopic body (the apparatus). Here it is, in principle, impossible to assume that the effect of the apparatus on the microparticle is small and unimportant. Let us consider a simple example. We assume that electrons are let successively through a slit in a screen. The screen with the slit is a macroscopic body (the apparatus) which measures the y -coordinate of an electron with an accuracy Δy , where Δy is the width of the slit.

The state of all the electrons before the interaction was the same. Let, for example, electrons with a definite direction of the momentum \mathbf{p} (e.g. along the x -axis) be incident on the apparatus. Here $p_y = 0$. The state of the apparatus before the interaction is also defined, but in a macroscopic way. In the process of interaction of the apparatus with the electron the latter is localized in the region Δy defined by the size of the opening in the screen. Then the state of the electron essentially changes. The electron passes from a state with the definite momentum component $p_y = 0$ to a state in which the momentum component p_y has a value lying in the interval $\Delta p_y \sim \hbar/\Delta y$. Indeed, as we know, diffraction occurs when electrons pass through a slit and the electrons get a momentum component along the y -axis. If we let electrons successively through the slit and measure the values of their momentum component p_y , then we shall obtain all possible values of p_y which lie in the interval Δp_y .

Thus we see that the effect of the apparatus on the electron changes the state of the latter and in principle cannot be made small. Although before the measurement the micro-object and the apparatus were in a definite state, the result of the interaction with the apparatus is not single-valued: we obtain a state with an indefinite value of the momentum component p_y . We can only find the probability of any one value of this quantity.

As a result of carrying out a successive series of measurements, no matter how large, we would not obtain a more accurate value of p_y , but only a more accurate expression for the distribution of the probabilities of different values of this quantity. If the micro-object were in given external conditions, then it would nevertheless be impossible to predict accurately the result of the measurement. One can speak only of the probability distribution of the results of the measurements. This is not associated with any shortcomings of the theory but with the very nature of microparticles. Hence it follows that the principle of mechanical determinism does not characterize the properties of microparticles.

A given initial value of a certain quantity and a definite law of interaction do not unambiguously determine the measured value of this quantity for the microparticle at subsequent instants of time. Thus the behaviour of an individual microparticle, and not just the behaviour of a set of microparticles, is determined by laws of a statistical type. The law of causality for a microparticle takes the following character. Let the state of a particle be known at the initial instant of time $t = 0$. This means that its wave function $\psi(\mathbf{r}, 0)$ is known. If all interactions which the microparticle undergoes are known, then, as we shall see below (see §6), its wave function can be determined unambiguously at subsequent instants of time $t > 0$. It follows from the meaning of the wave function that we can by this predict the probabilities (see §21) that the quantities characterizing the particle (the coordinate, momentum, energy and others) will have particular values at any instant of time $t > 0$.

The principle of causality formulated in quantum mechanics in this way is of considerably more general character than the dynamical regularity (the Laplace determinism) of classical mechanics*.

* The reader may find in the work of V.A. Fok a more detailed consideration of the problems which are touched upon in this section. The interpretation of quantum mechanics may be found in the collection of papers *Filozoficheskie voprosy sovremenoi fiziki* (*Philosophic problems of contemporary physics*) published by the Academy of Sciences of the USSR in 1959. See also N. Bohr, *Atomic physics and human knowledge* (Wiley, New York, 1958).

The Schrödinger Equation

§6. The Schrödinger wave equation

In §2 we established the form of the wave function describing the motion of a free particle with a given value of the momentum. This wave function had the form of a plane de Broglie wave. We now turn to the consideration of the motion of particles in external fields of force. For this it is necessary to find the wave function describing the motion of a particle in a given field of force. It turns out that it is possible to establish the form of the differential equation satisfied by the wave function. One can find the wave function itself from the solution of this equation. It should be noted, first of all, that the equation for the wave function must be linear. Indeed, functions satisfying non-linear equations obviously do not meet the requirements of the principle of superposition. Further, it is clear that the wave function we already know, which describes the motion of a free particle, must be the solution of the required differential equation in the particular case where the field is absent. Finding the linear differential equation satisfied by a plane de Broglie wave

$$\psi(x, t) = A e^{(i/\hbar)(\mathbf{p} \cdot \mathbf{r} - Et)}, \quad (6.1)$$

presents no difficulty. For this we note that

$$\frac{\partial \psi}{\partial t} = -\frac{i}{\hbar} E \psi.$$

Furthermore,

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = -\frac{1}{\hbar^2} (p_x^2 + p_y^2 + p_z^2) \psi.$$

Taking into account that for a free particle

$$\frac{p_x^2 + p_y^2 + p_z^2}{2m} = E, \quad (6.2)$$

we find

$$\frac{\partial \psi}{\partial t} = \frac{i\hbar}{2m} \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right).$$

This equation is usually written in the form

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi. \quad (6.3)$$

This linear differential equation in partial derivatives is called the Schrödinger equation. It does not contain any characteristics of the state of the particle, for example the value of its momentum or energy. It involves only the mass of the particle as well as the universal constant \hbar . Equation (6.3) is evidently satisfied not only by a wave function of the form (6.1), which represents the wave function of a particle with a given value of the momentum, but also by any superposition of such wave functions.

The Schrödinger equation possesses the feature that it is an equation of the first order in time and contains the factor i . The latter means that the wave function must be complex.

We note that it seems that a function expressed by a real relation, for example in the form of a travelling wave $\psi = A \cos \hbar^{-1}(\mathbf{p} \cdot \mathbf{r} - Et)$, could be chosen as the wave function of a free particle. However, we could not then construct any equation of the first order in time whose solution would be an arbitrary superposition of such functions. The fact that the Schrödinger equation contains only the first derivative of the wave function with respect to time is closely associated with the expression of the principle of causality in quantum mechanics (see §5). Indeed, if the Schrödinger equation contained, for example, the second derivative of the wave function with respect to time, then the knowledge of the wave function at the initial instant of time would

be insufficient for determining the wave function at an arbitrary instant of time t . Namely, it would also be necessary to give the value of the first derivative of the wave function with respect to time for the initial instant of time.

Among the solutions of eq. (6.3) there are solutions depending harmonically on the time

$$\psi(x, t) = \psi(x) e^{-(i/\hbar)Et}. \quad (6.4)$$

Substituting (6.4) into (6.3), we obtain an equation for a function which depends only on the coordinates of the particle

$$\nabla^2 \psi(x) + \frac{2mE}{\hbar^2} \psi(x) = 0. \quad (6.5)$$

This equation defines the function $\psi(x)$ for a free particle. Let us generalize eq. (6.5) to the case of a particle moving in a field of force. This generalization is based on the following assumption: the energy E involved in eq. (6.5) represents the kinetic energy of the particle. Indeed, for free motion the kinetic energy is the same as the total energy. If in the required generalization the energy E occurring in eq. (6.5) is assumed to be the total energy, then the wave function describing the motion of electrons in a field of force will not depend on the forces acting on the particle. However, this would make no sense. Thus we arrive at the conclusion that E in eq. (6.5) must be understood to be the kinetic energy of the particle. Denoting the potential energy of the particle by $U(x)$, and the total energy by E , we get

$$\nabla^2 \psi(x) + \frac{2m}{\hbar^2} [E - U(x)] \psi(x) = 0. \quad (6.6)$$

Equation (6.6) represents the required generalization of the Schrödinger equation to the case of a particle moving in an arbitrary potential field which does not depend on time. This equation only determines the dependence of the wave function on the coordinates, while the dependence on time is determined as before by the relation (6.4).

Equation (6.6) is called the Schrödinger equation for stationary states. Indeed, the probability density of the measurement of the coordinates of a particle in a state (6.4) does not depend on the time

$$|\psi(x, t)|^2 = |\psi(x, 0)|^2. \quad (6.7)$$

In §28 it will be shown that the probabilities of the measurement of other physical quantities in a state (6.4) also do not depend on the time.

Substituting the derivative with respect to time $\partial\psi/\partial t$ for the quantity $E\psi$

by means of (6.4), we arrive at the general Schrödinger wave equation

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + U\psi, \quad (6.8)$$

where the wave function ψ depends on the coordinates x, y, z and the time t .

Equation (6.8) is the fundamental equation of quantum mechanics. It plays in quantum mechanics the same role as Newton's equations in classical mechanics, and could be called the equation of motion of a quantum particle. To define the law of motion of a particle in quantum mechanics means to define the value of the ψ -function at every instant of time and at every point of space.

It should be noted that the above reasoning is not a derivation of the Schrödinger equation in the strict sense of the word. Like Newton's and Maxwell's equations, the Schrödinger equation appeared, on the one hand, as the generalization of known experimental data and, on the other hand, as a great scientific prediction.

We shall see later how the discreteness of energy levels follows from the Schrödinger equation. Also it will become clear that the Schrödinger equation satisfies the correspondence principle. The validity of the Schrödinger equation and the interpretation of the meaning of the wave function involved in it are confirmed by a vast amount of experimental data from contemporary atomic and nuclear physics. In order to obtain the law of motion of a particle, the wave function $\psi(x, t)$, the initial and boundary conditions must be given in addition to the Schrödinger equation. Since the Schrödinger equation is an equation of the first order in time, it is necessary to know the initial value of the wave function $\psi(x, 0)$.

The set of boundary conditions in general amounts to the requirement of single-valuedness and continuity of the wave function and its first derivatives, and to the fulfillment of certain normalization conditions. The latter is usually a bounding condition on the modulus of the wave function. The whole set of the initial condition and the conditions of single-valuedness, continuity and finiteness of the wave function and of its first derivatives makes it possible, in principle, to find a unique solution of the Schrödinger equation: the wave function $\psi(x, t)$. In other words, if the initial value of the wave function is given, then from the solution of the Schrödinger equation one can determine unambiguously the state of the quantum system for any subsequent instant of time $t > 0$. Namely, for $t > 0$ one can find the wave function of the system $\psi(x, t)$.

We shall see in §23 that by defining $\psi(x, t)$ the quantum particle is characterized as fully as, for example, the particle in classical mechanics is by defin-

ing its trajectory. We note in addition that in certain problems of quantum mechanics it is convenient to approximate the potential energy by a discontinuous function. At the point of discontinuity of the potential energy the wave function and its first derivatives must remain continuous. The derivative of the wave function undergoes a jump only at the surface of an infinitely large discontinuity of the potential energy.

The Schrödinger equation, as well as the equations of motion of classical mechanics, allows 'reversal in time'.

Indeed, eq. (6.8) does not change when the transformation $t \rightarrow -t$ and the transition to the complex conjugate function ψ^* are made. Consequently, a process reversed in time is described by the wave function $\psi_{\text{rev}}(x, t)$

$$\psi_{\text{rev}}(x, t) = \psi^*(x, -t). \quad (6.9)$$

We note that for motion in a magnetic field reversal in time takes place only when the direction of the magnetic field is also reversed (see §27). We shall consider the problem of time reversal in more detail in §98.

§7. The probability current density

Generally speaking the wave function describing the motion of a particle changes in space and time. However, this change cannot be arbitrary.

That is, a conservation law holds. To formulate this law, we consider the integral $\int_V |\psi|^2 dV$, which represents the probability of finding the particle in the volume V . Proceeding in the same way as in deriving the law of charge conservation (see §5 of Part I), we find the derivative of the above integral with respect to time. To calculate $\partial\psi/\partial t$ and $\partial\psi^*/\partial t$ we make use of the Schrödinger equation (6.8) and its conjugate equation. Then we obtain

$$\begin{aligned} \frac{\partial}{\partial t} \int \psi \psi^* dV &= \int \left(\frac{\partial \psi}{\partial t} \psi^* + \psi \frac{\partial \psi^*}{\partial t} \right) dV = \\ &= \frac{\hbar}{2mi} \int (\psi \nabla^2 \psi^* - \psi^* \nabla^2 \psi) dV = \\ &= \frac{\hbar}{2mi} \int \nabla \cdot (\psi \nabla \psi^* - \psi^* \nabla \psi) dV. \end{aligned} \quad (7.1)$$

Making use of the Gauss-Ostrogradsky theorem, we have

$$\int_V \nabla \cdot (\psi \nabla \psi^* - \psi^* \nabla \psi) dV = \oint_S (\psi \nabla \psi^* - \psi^* \nabla \psi) dS,$$

where the surface S bounds the volume V . Hence

$$\frac{\partial}{\partial t} \int_V |\psi|^2 dV = \frac{\hbar}{2mi} \oint_S (\psi \nabla \psi^* - \psi^* \nabla \psi) dS. \quad (7.2)$$

We introduce the vector \mathbf{j} defined by the relation

$$\mathbf{j} = \frac{\hbar}{2mi} (\psi^* \nabla \psi - \psi \nabla \psi^*). \quad (7.3)$$

Then (7.2) is rewritten in the form

$$-\frac{\partial}{\partial t} \int_V |\psi|^2 dV = \oint_S j_n dS. \quad (7.4)$$

Formula (7.4) shows that the probability density satisfies a conservation law, and the vector \mathbf{j} we introduced has the meaning of the probability current density. Relation (7.4) can be rewritten in differential form as the continuity equation

$$\frac{\partial |\psi|^2}{\partial t} + \nabla \cdot \mathbf{j} = 0. \quad (7.5)$$

The integral of the normal component of the vector \mathbf{j} with respect to a particular surface represents the probability that the particle will cross the surface mentioned in unit time.

Let us consider, in particular, free motion. We take the wave function in the form of the plane wave $\psi = A e^{(i/\hbar)(\mathbf{p} \cdot \mathbf{r} - Et)}$. Making use of the relation (7.3), we obtain

$$\mathbf{j} = \frac{1}{m} \mathbf{p} |A|^2. \quad (7.6)$$

We now apply the relation (7.4) to all space, i.e. we assume the surface S to be infinitely distant. If ψ is a quadratically integrable function, then the integrand in the integral with respect to the surface decreases more rapidly than r^{-4} , and the surface of integration increases in proportion to r^2 . As a result the integral over the surface in (7.4) reduces to zero. But if ψ does not tend to zero in the way mentioned when $r \rightarrow \infty$, as, for example, in the case of a plane wave, then there is a current of particles at infinity. If this current is stationary, then the wave function can be normalized in such a way that the vector \mathbf{j} is the particle-current density vector.

Finally, we note from formula (7.3) that the current density \mathbf{j} obviously

reduces to zero if the state of the system is described by a real wave function ψ .

The relation (7.5) written in the form

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0, \quad (7.7)$$

can be interpreted as the law of conservation of the number of particles (see §5 of Part I).

§8. A particle in a one-dimensional rectangular potential well

Before turning to the consideration of real atomic systems we shall discuss the general properties of the solutions of the Schrödinger equation by the use of some simple models. Let us consider first of all the one-dimensional motion of a particle in a potential field defined as follows:

$$U(x) = \begin{cases} 0 & \text{for } 0 < x < l, \\ \infty & \text{for } x \leq 0 \text{ and } x \geq l. \end{cases}$$

We shall call such a potential field an infinitely deep potential well. It is clear that in such a well a particle can move only in the region of space $0 \leq x \leq l$.

At the boundary of the well the particle is acted upon by arbitrarily large forces which prevent it from getting out, so that the particle behaves as if confined to a region of space bounded by perfectly reflecting walls. It turns out that by such a simple example one can establish a number of properties of quantum-mechanical systems. It is important that these properties are not associated with the model but are of a general character. Furthermore, interest in this problem is also due to the fact that a model of a potential well is often successfully used for a rough description of a number of systems, for example electrons in a metal or nucleons in a nucleus.

The solution of the Schrödinger equation must be written for two regions: outside the potential well and inside it. Since the particle cannot be outside the potential well, its wave function is equal to zero outside the interval $0 \leq x \leq l$. From the condition of continuity it follows that the wave function is also equal to zero at the points $x = 0$ and $x = l$, i.e. that

$$\psi(0) = \psi(l) = 0. \quad (8.1)$$

The requirement (8.1) is the boundary condition for the solution of the Schrödinger equation inside the potential well. In the region $0 < x < l$ the

Schrödinger equation for stationary states (6.6) is of the form

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} = E\psi. \quad (8.2)$$

The solution of this equation can evidently be written as

$$\psi = A \sin(kx + \alpha), \quad (8.3)$$

where $k = (2mE/\hbar^2)^{1/2}$. We now make use of the boundary conditions (8.1). From the relation $\psi = 0$ for $x = 0$ it follows that $\alpha = 0$. The condition $\psi(l) = 0$ gives

$$kl = n\pi, \quad (8.4)$$

where n is any integer larger than zero. For $n = 0$ we would have $\psi \equiv 0$, which would mean the absence of the particle in all space. The condition (8.4) makes it possible to find the possible values of the energy of the particle

$$E_n = \frac{\pi^2 \hbar^2}{2ml^2} n^2. \quad (8.5)$$

We see that the Schrödinger equation has solutions satisfying the boundary conditions only for discrete values of the quantum number n . Thus the energy of the particle in the infinitely deep potential well turns out to be quantized. The discreteness of the energy arose in a natural way, without any subsidiary assumptions. In the case given it turned out to follow directly from the boundary conditions imposed upon the wave function at the limits of the integration range. The state of the particle which has the lowest possible energy will henceforth be called the normal or ground state, while all other states will be called excited states. The energy of the ground state of a particle in an infinitely deep potential well is obtained from formula (8.5) for $n = 1$:

$$E_1 = \frac{\pi^2 \hbar^2}{2ml^2}. \quad (8.6)$$

We note that the minimum energy value of the particle is consistent with the uncertainty principle. Indeed, the uncertainty in the coordinate of the particle is $\Delta x \sim l$. The uncertainty in the momentum, Δp , is of the order of \hbar/l . Since $p \geq \Delta p$, the minimum energy of the particle turns out to be equal to

$$\frac{p^2}{2m} \gtrsim \frac{\hbar^2}{2ml^2},$$

which is in order of magnitude the same as (8.6).

Let us now determine the spacing between neighbouring energy levels

($\Delta n = 1$):

$$\Delta E_n = E_{n+1} - E_n = \frac{\pi^2 \hbar^2}{2ml^2} (2n + 1).$$

The spacing between levels increases with decreasing mass of the particle and size of the region of its motion l . Thus, for example, for an electron ($m \sim 10^{-27}$ g) confined in a region $l \sim 5 \times 10^{-8}$ cm we find $\Delta E \sim 1$ eV. On the contrary, in the case of a molecule with $m \sim 10^{-23}$ g moving, for example, in a region $l \sim 10$ cm, the spacing between levels amounts to $\Delta E \sim 10^{-20}$ eV. This spacing is so small, for example in comparison with $kT = 0.025$ eV, that in practice the energy of a molecule can be considered to be a continuously varying quantity.

Let us find the ratio $\Delta E_n/E_n$, i.e. the relative spacing between energy levels. We see that $\Delta E_n/E_n \sim n^{-1}$ and tends to zero for very large n . The discreteness of quantum states is no longer significant for large quantum numbers and in fact a transition to a quasi-continuous variation of the energy takes place.

Let us consider in somewhat more detail the properties of the wave function of a particle in a potential well. The wave function corresponding to the n th energy level is of the form

$$\psi_n = A_n \sin \frac{\pi n}{l} x. \quad (8.7)$$

We define the constant A_n by the normalization condition

$$\int_0^l |\psi_n|^2 dx = 1.$$

Then

$$|A_n|^2 \int_0^l \sin^2 \frac{n\pi}{l} x dx = |A_n|^2 \int_0^l \frac{1}{2} \left(1 - \cos \frac{2n\pi}{l} x \right) dx = |A_n|^2 \frac{l}{2} = 1.$$

Hence

$$A_n = (2/l)^{1/2}. \quad (8.8)$$

Thus the value of the constant does not depend on the quantum number n .

The probability density $|\psi|^2$ of finding the particle at different points inside the well is illustrated in fig. V.2. In classical mechanics a particle moving in a potential well can be found with equal probability at any point

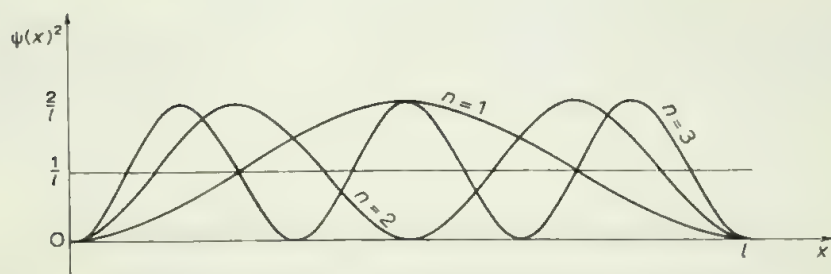


Fig. V.2

inside the well (straight line in fig. V.2). Indeed, the probability dW_{class} of observing the particle in an interval dx is proportional to the time dt of the particle's being in this interval:

$$dW_{\text{class}} \sim dt = \frac{1}{v} dx.$$

Since a particle inside a well is not acted upon by any forces it moves with constant velocity v and, consequently, dW_{class} does not depend on x . As the quantum number n (the energy of the particle) increases the maxima of the probability distribution tend to approach each other. In the limit $n \rightarrow \infty$ the probability distribution obtained from the quantum-mechanical calculation leads to the same results as the classical distribution. This follows from the fact that the function $\sin^2(\pi nx/l)$ rapidly oscillates as x changes, and in integration over any finite interval can be replaced by $\frac{1}{2}$. Thus consideration of the simplest quantum-mechanical system leads us to the following conclusions which, as we shall see later, are of a general character:

- (1) the energy of a microparticle moving in a potential well runs over a discrete sequence of values;
- (2) even for $E = E_1$ (ground state) the particle is not in a state of complete rest with kinetic energy equal to zero;
- (3) the discrete character of the energy levels manifests itself when the mass of the particle and the size of the region in which the motion takes place are small;
- (4) for large values of the quantum numbers the quantum-mechanical relations go over into the formulae of classical physics. This statement is a particular case of the correspondence principle which we shall frequently encounter.

Later, in considering the quantum oscillator or atomic systems, we shall

see that quantization of states can take place even in systems which are not confined by impenetrable walls. At the same time we shall see that the presence of discrete energy states is not a necessity for quantum-mechanical systems. In certain cases quantum-mechanical systems have a continuous energy spectrum.

§9. A particle in a three-dimensional rectangular potential well

Let us now consider the more complex case of the motion of a particle in a three-dimensional infinitely deep potential well. We shall assume that the region of space in which the particle moves is defined by the inequalities $0 < x < l_1$, $0 < y < l_2$ and $0 < z < l_3$. In this case the wave equation can be written in the form

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) = E\psi. \quad (9.1)$$

The boundary conditions are analogous to (8.1) and have the form

$$\begin{aligned} \psi(0, y, z) &= \psi(x, 0, z) = \psi(x, y, 0) = \\ &= \psi(l_1, y, z) = \psi(x, l_2, z) = \psi(x, y, l_3) = 0. \end{aligned} \quad (9.2)$$

We write the solution of eq. (9.1) as follows:

$$\psi = B \sin k_1 x \sin k_2 y \sin k_3 z. \quad (9.3)$$

Substituting ψ into the equation, we obtain the relation

$$\frac{\hbar^2}{2m} (k_1^2 + k_2^2 + k_3^2) = E. \quad (9.4)$$

From the boundary conditions (9.2) it follows that

$$k_1 l_1 = n_1 \pi, \quad k_2 l_2 = n_2 \pi, \quad k_3 l_3 = n_3 \pi, \quad (9.5)$$

where n_1, n_2 and n_3 are integers.

Substituting the values of k_1, k_2 and k_3 into (9.4) and (9.3), we obtain the expressions for the energy and for the wave function

$$E_{n_1, n_2, n_3} = \frac{\pi^2 \hbar^2}{2m} \left(\frac{n_1^2}{l_1^2} + \frac{n_2^2}{l_2^2} + \frac{n_3^2}{l_3^2} \right), \quad (9.6)$$

$$\psi_{n_1 n_2 n_3} = B \sin \frac{\pi n_1 x}{l_1} \sin \frac{\pi n_2 y}{l_2} \sin \frac{\pi n_3 z}{l_3}. \quad (9.7)$$

The constant B is again defined by the normalization condition

$$\int_V |\psi_{n_1 n_2 n_3}|^2 dx dy dz = 1$$

and is equal to

$$B = \left(\frac{8}{l_1 l_2 l_3} \right)^{\frac{1}{2}}. \quad (9.8)$$

Let us consider, in particular, a particle moving in a potential well of cubic form, i.e. the case $l_1 = l_2 = l_3 = l$. The energy of the particle is equal to

$$E_{n_1, n_2, n_3} = \frac{\pi^2 \hbar^2}{2ml^2} (n_1^2 + n_2^2 + n_3^2). \quad (9.9)$$

From formula (9.9) it is easily seen that one and the same energy value can be realized by means of different combinations of the numbers n_1 , n_2 and n_3 . This means that several different quantum states with different wave functions correspond to one and the same energy value. Such energy levels are said to be degenerate, and the number of different states corresponding to a given energy level is called the multiplicity of degeneracy.

Let us consider, for example, the energy level

$$E = \frac{\pi^2 \hbar^2}{2ml^2} \cdot 6,$$

where $n_1^2 + n_2^2 + n_3^2 = 6$. Since each of the n 's is an integer larger than zero, this equality can be satisfied by the three different combinations of the numbers n_1, n_2, n_3

- (1) $n_1 = 2, \quad n_2 = 1, \quad n_3 = 1,$
- (2) $n_1 = 1, \quad n_2 = 2, \quad n_3 = 1,$
- (3) $n_1 = 1, \quad n_2 = 1, \quad n_3 = 2.$

Thus to the given energy level there correspond 3 different states ψ_{211} , ψ_{121} and ψ_{112} . Consequently, the multiplicity of degeneracy is equal to three. In considering more complex systems, for example atoms, we shall frequently encounter the phenomenon of degeneracy.

§ 10. The quantum-mechanical oscillator

Turning to more complex quantum-mechanical systems, we shall consider

the theory of the linear harmonic oscillator. Such an oscillator represents the quantum analogue of a particle performing small linear oscillations about an equilibrium position. An example of small oscillations in atomic systems are the small oscillations of atoms in a molecule (see §41 of Part III).

Another no less important example is the thermal motion of a crystal, which amounts to a set of linear harmonic oscillators. We shall also deal with the problem of the harmonic oscillator in quantum electrodynamics, where an arbitrary electromagnetic field is represented in the form of the superposition of independent quantum oscillators (see §101).

The above examples show that the theory of the linear harmonic oscillator is one of the important problems of quantum mechanics.

The potential energy of a linear harmonic oscillator is given by the well-known formula $U = \frac{1}{2}m\omega^2 x^2$. Hence the Schrödinger equation (6.6) for the linear harmonic oscillator has the form

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + \frac{m\omega^2 x^2}{2} \psi = E\psi. \quad (10.1)$$

In solving it, it is convenient to go over to dimensionless variables

$$\xi = \left(\frac{m\omega}{\hbar}\right)^{\frac{1}{2}} x; \quad \lambda = \frac{2E}{\hbar\omega}. \quad (10.2)$$

In the new notation the Schrödinger equation takes the form

$$-\frac{d^2\psi}{d\xi^2} + \xi^2\psi = \lambda\psi. \quad (10.3)$$

An important difference of the oscillator from the examples considered before is the fact that in this case the motion of the particle is not restricted by an impenetrable wall. Hence the oscillator has no boundary conditions similar to the conditions (8.1). The only requirement imposed upon the wave function of the oscillator is the requirement that it should be quadratically integrable. We shall see that the Schrödinger equation for the oscillator has a solution satisfying this requirement only for certain definite values of the parameter λ . These values are called the eigenvalues of eq. (10.3).

In order to explain the general character of the solutions of this equation, let us consider the asymptotic behaviour of $\psi(\xi)$ for very large values of the argument $\xi \gg \lambda$.

For $\xi \gg \lambda$ in eq. (10.3) one can drop $\lambda\psi$ as compared with $\xi^2\psi$. We then have, obviously,

$$\frac{d^2\psi}{d\xi^2} - \xi^2\psi = 0. \quad (10.4)$$

The solution of this equation satisfying the requirement of finiteness for large ξ is the function

$$\psi = A \xi^m e^{-\frac{1}{2}\xi^2}, \quad (10.5)$$

where A is a constant, and m is an arbitrary finite number.

The second independent solution of eq. (10.4), $\psi \sim e^{+\frac{1}{2}\xi^2}$, increases indefinitely for $\xi \rightarrow \infty$ and must be dropped.

We try to find a solution of eq. (10.3) in the form

$$\psi = e^{-\frac{1}{2}\xi^2} f(\xi), \quad (10.6)$$

where $f(\xi)$ is a new unknown function which for $\xi \rightarrow \infty$ behaves as ξ^m . Substituting (10.6) into (10.3), we arrive at the following equation for the function f :

$$\frac{d^2 f}{d\xi^2} - 2\xi \frac{df}{d\xi} + (\lambda - 1)f = 0. \quad (10.7)$$

Since the point $\xi = 0$ is not a singular point of eq. (10.7), we seek a solution of this equation in the form of a power series

$$f(\xi) = \sum_{k=0}^{\infty} a_k \xi^k \quad (10.8)$$

The derivatives $df/d\xi$ and $d^2 f/d\xi^2$ have the form

$$\frac{df}{d\xi} = \sum k a_k \xi^{k-1}, \quad \frac{d^2 f}{d\xi^2} = \sum k(k-1) a_k \xi^{k-2}. \quad (10.9)$$

Substituting the series (10.9) into eq. (10.7), we obtain

$$\sum k(k-1) a_k \xi^{k-2} - 2\xi \sum k a_k \xi^{k-1} + (\lambda - 1) \sum a_k \xi^k = 0. \quad (10.10)$$

In order that a power series of the form $\sum_n c_n \xi^n$ be identically equal to zero, it is necessary that all coefficients c_n reduce to zero. Assuming the coefficient of ξ^k to be equal to zero, we obtain the recurrence formula

$$a_{k+2} = \frac{2k+1-\lambda}{(k+2)(k+1)} a_k. \quad (10.11)$$

It is easily seen that for $\xi \rightarrow \infty$ such a series behaves as e^{ξ^2} , since in this case large k are essential and (10.11) gives $a_{k+2} \approx (2/k)a_k$. In this case the function ψ of (10.6) increases indefinitely. Such a solution must be eliminated.

We obtain a solution satisfying the necessary conditions of finiteness and behaving for $\xi \rightarrow \infty$ as (10.5) only in the case where the series (10.8) reduces

to a polynomial, i.e. is cut off at a certain term. Thus suppose that $a_n \neq 0$, $a_{n+2} = 0$. Then all subsequent coefficients also vanish, and the function f reduces to a polynomial of the n th degree.

It follows from (10.11) that in this case the condition

$$2n + 1 - \lambda = 0 \quad (10.12)$$

is necessary, where n is an integer, $n \geq 0$, since n is the ordinal number of the term at which the series ends.

Substituting this value of λ into (10.2), we obtain

$$E_n = \hbar\omega(n + \frac{1}{2}). \quad (10.13)$$

Hence it is seen that the energy of the oscillator can only take on discrete values, and the energy levels are equally spaced at intervals of $\hbar\omega$.

We write the wave function corresponding to the n th excited energy level in the form

$$\psi_n(\xi) = A_n e^{-\frac{1}{2}\xi^2} f_n(\xi), \quad (10.14)$$

where $f_n(\xi)$ is a polynomial of the n th power with coefficients which are defined by the relation (10.11), and A_n is a factor determined by the normalization condition. The polynomials $f_n(\xi)$ are called the Chebyshev-Hermite polynomials and are denoted by $H_n(\xi)$. The Chebyshev-Hermite polynomials are often written in the form

$$H_n(\xi) = (-1)^n e^{\xi^2} \frac{d^n e^{-\xi^2}}{d\xi^n}. \quad (10.15)$$

They satisfy the differential equation

$$\frac{d^2 H_n}{d\xi^2} - 2\xi \frac{dH_n}{d\xi} + 2nH_n = 0, \quad (10.16)$$

which is obtained from (10.7) taking into account condition (10.12).

We give the first four Chebyshev-Hermite polynomials:

$$\begin{aligned} H_0(\xi) &= 1, & H_1(\xi) &= 2\xi, & H_2(\xi) &= 4\xi^2 - 2, \\ H_3(\xi) &= 8\xi^3 - 12\xi, & H_4(\xi) &= 16\xi^4 - 48\xi^2 + 12. \end{aligned} \quad (10.17)$$

Knowing the general form of the Chebyshev-Hermite polynomials, one can

calculate the normalization integral. One then obtains* for A_n

$$A_n = \left(\frac{m\omega}{\hbar\pi} \right)^{\frac{1}{4}} \left(\frac{1}{n!2^n} \right)^{\frac{1}{2}}. \quad (10.18)$$

The form of the wave functions for different quantum numbers n is shown in fig. V.3. We note that the wave function corresponding to the ground state $n = 0$ of the oscillator is nowhere equal to zero. The wave function corresponding to the level $n = 1$ is equal to zero once, at $x = 0$, while $\psi_2(x)$ ($n = 2$) is equal to zero twice, and so on. The points at which the wave function is equal to zero are called the nodes of the wave function. It is easily seen that the number of nodes of the wave function is equal to the quantum number n .

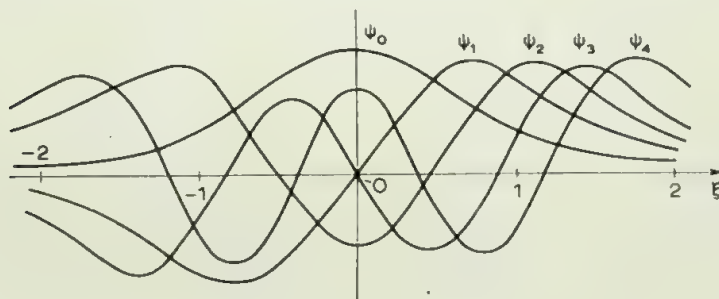


Fig. V.3

This statement is not specific to the oscillator. It can be stated** that in general in the one-dimensional case the number of nodes of the wave function is determined by the quantum number n . The probability of finding the particle at the point x in the interval dx is equal to

$$W_n(x) dx = |\psi_n(x)|^2 dx.$$

These probabilities for different n are shown in fig. V.4. Let us compare the expressions obtained with the probability of finding the particle at a given point as calculated by means of classical mechanics. The latter is defined as the ratio of the time dt spent in the neighbourhood of the given point to the period of the motion. The classical probability turns out to be highest in the

* See, for example, L.D.Landau and E.M.Lifshitz, *Quantum mechanics, Non-relativistic theory* (Pergamon Press, Oxford, 1965).

** R.Courant and D.Hilbert, *Methods of mathematical physics*, Vol. 1 (Interscience, New York, 1953).

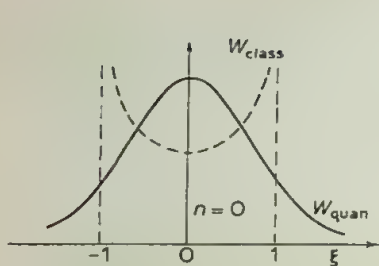


Fig. V.4a

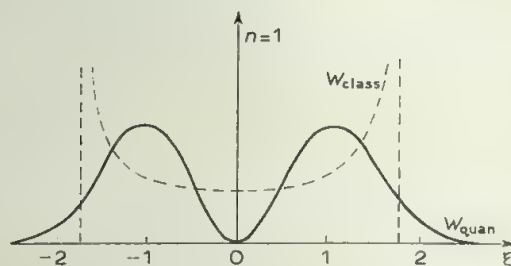


Fig. V.4b

neighbourhood of the turning points $x = \pm x_0$, at which the velocity of motion reduces to zero. On the other hand, in the neighbourhood of the point $x = 0$ the particle has its largest velocity and the probability of finding it is a minimum.

It is seen from the curves of fig. V.4 that the probability of finding a quantum particle differs from zero even in the region outside the turning points, which is unattainable in classical mechanics. For large quantum numbers (fig. V.5) the quantum probability distribution approaches the classical one, in agreement with the correspondence principle.

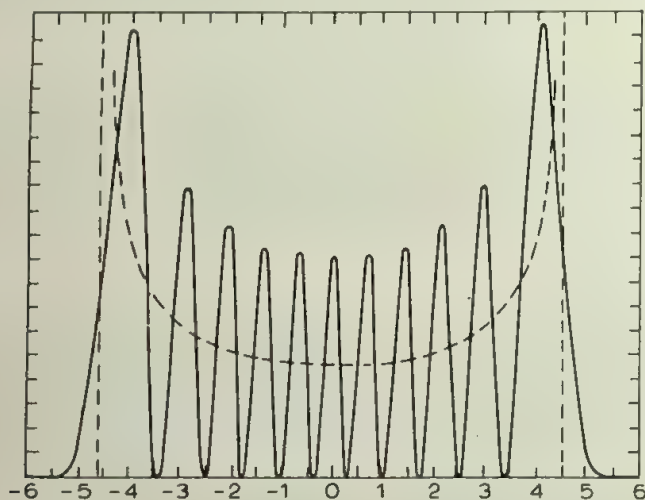


Fig. V.5

In conclusion we note that the lowest possible energy value of the oscillator, equal to $\frac{1}{2}\hbar\omega$, is different from zero. This means that the quantum oscillator can never be in a state of absolute rest. This fact in its turn is associated with the uncertainty principle. In order of magnitude the energy of the oscillator is

$$E \geq \frac{\Delta p^2}{2m} + \frac{m\omega^2}{2} \Delta x^2 \geq \frac{\Delta p^2}{2m} + \frac{m\omega^2}{2} \left(\frac{\hbar}{\Delta p} \right)^2.$$

Considering this quantity as a function of Δp , it is easily established that it has a minimum for $\Delta p \sim (m\omega\hbar)^{\frac{1}{2}}$ and is of the order of magnitude of $\hbar\omega$. Experimentally the zero-point energy E_0 is observed in the scattering of light by a crystal at a temperature close to absolute zero. At absolute zero the crystal is in the ground (lowest energy) state. Nevertheless, atoms perform zero-point oscillations which cause scattering of the light.

§ 11. The three-dimensional oscillator

Let us now consider the motion of a spatial three-dimensional oscillator. For generality we shall assume that the natural frequencies are different in three mutually perpendicular directions and equal respectively to ω_1 , ω_2 and ω_3 . Then the potential energy is expressed by the formula

$$U = \frac{m\omega_1^2}{2}x^2 + \frac{m\omega_2^2}{2}y^2 + \frac{m\omega_3^2}{2}z^2. \quad (11.1)$$

The Schrödinger equation correspondingly has the form

$$-\frac{\hbar^2}{2m} \nabla^2 \psi + \frac{m}{2} (\omega_1^2 x^2 + \omega_2^2 y^2 + \omega_3^2 z^2) \psi = E \psi. \quad (11.2)$$

We try to find the solution of eq. (11.2) in the form of a product of functions each of which depends on only one coordinate

$$\psi(x, y, z) = \psi_1(x) \psi_2(y) \psi_3(z). \quad (11.3)$$

Substituting (11.3) into (11.2) and separating the variables we obtain

$$-\frac{\hbar^2}{2m} \frac{d^2 \psi_i}{dx_i^2} + \frac{m\omega_i^2 x_i^2}{2} \psi_i(x_i) = E_i \psi_i(x_i), \quad i = 1, 2, 3, \quad (11.4)$$

where $x_1 = x$, $x_2 = y$, $x_3 = z$; $E_1 + E_2 + E_3 = E$.

Thus the problem is reduced to the one-dimensional case. In correspon-

dence with this, making use of (10.13), (10.14) and (10.18) we can write

$$\psi_{n_1 n_2 n_3}(x_1, x_2, x_3) = \left(\frac{m^3 \omega_1 \omega_2 \omega_3}{h^3 \pi^3} \right)^{\frac{1}{4}} \left(\frac{2^{-(n_1+n_2+n_3)}}{n_1! n_2! n_3!} \right)^{\frac{1}{2}} \times \\ \times \exp[-\frac{1}{2}(\xi_1^2 + \xi_2^2 + \xi_3^2)] H_{n_1}(\xi_1) H_{n_2}(\xi_2) H_{n_3}(\xi_3), \quad (11.5)$$

where $\xi_i = (m\omega_i/h)^{\frac{1}{2}} x_i$ ($i = 1, 2, 3$).

The total energy of the oscillator is equal to

$$E = \hbar\omega_1(n_1 + \frac{1}{2}) + \hbar\omega_2(n_2 + \frac{1}{2}) + \hbar\omega_3(n_3 + \frac{1}{2}). \quad (11.6)$$

In particular, for an isotropic oscillator, which has $\omega_1 = \omega_2 = \omega_3 = \omega$, the total energy is of the form

$$E_n = \hbar\omega(n + \frac{3}{2}), \quad \text{where } n = n_1 + n_2 + n_3, \quad (11.7)$$

i.e. E_n depends on the sum of the quantum numbers n_1 , n_2 and n_3 . This means that a given energy value (given n) can be obtained from different combinations of n_1 , n_2 and n_3 . Hence it follows that all the energy levels, except the ground level $n = 0$, are degenerate. The multiplicity of degeneracy is easily calculated. For this we fix, in addition to n , the quantum number n_1 . Then the number of possible combinations of n_1 , n_2 and n_3 will be equal to the number of possible values of n_2 , i.e. equal to $n - n_1 + 1$, since n_2 can vary from zero up to $n - n_1$. Summing the expression obtained over all possible values of the number n_1 , we find the total number of combinations of the three quantum numbers n_1 , n_2 and n_3 which add up to the given number n , i.e. the multiplicity of degeneracy of the n th energy level:

$$\sum_{n_1=0}^n (n - n_1 + 1) = \frac{1}{2}(n+1)(n+2). \quad (11.8)$$

§12. Reflection from and penetration through a potential barrier

Among the other relatively simple problems of quantum mechanics we shall consider the motion of particles in a field of force which has the form of a potential barrier. This means that the forces act on the particle in a certain limited region of space. Outside this region the particle moves as a free particle. We shall see that the study of the motion of particles in a field having the form of a barrier of the simplest shape will allow a number of important and, in principle new, properties of quantum particles to be exhibited. We shall

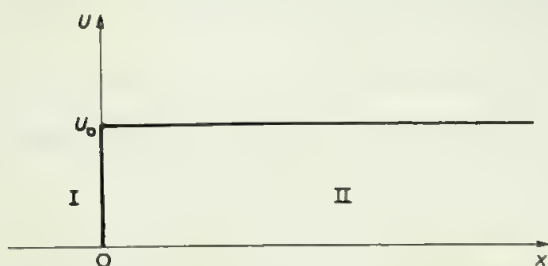


Fig. V.6

begin our consideration with the simple rectangular one-dimensional barrier of infinite extent shown in fig. V.6. In classical mechanics any particle moving from left to right with an energy smaller than the barrier height U_0 is completely reflected from the potential wall. The region $x > 0$ is inaccessible to it, since in this region the total energy of the particle would be less than the potential energy. This would mean that the kinetic energy must be negative, which is evidently impossible. If, on the contrary, E is larger than U_0 , then according to the laws of classical mechanics the particle passes freely above the barrier, moving in the region $x > 0$ with a lower kinetic energy equal to $E - U_0$.

Let us now consider the motion of a particle in the same situation according to the laws of quantum mechanics. For this we write the Schrödinger equation for the stationary states of the particle in the field of a barrier of infinite extent

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + U(x)\psi = E\psi, \quad (12.1)$$

where U is the potential energy whose graph is shown in fig. V.6. The solutions of eq. (12.1) are conveniently considered in two different regions. Region I ranges from $x = -\infty$ to $x = 0$, and region II from $x = 0$ to $x = \infty$.

We write the Schrödinger equation for each of the regions mentioned:

$$\begin{aligned} \frac{d^2\psi}{dx^2} + k^2\psi &= 0, & x < 0, \\ \frac{d^2\psi}{dx^2} + k'^2\psi &= 0, & x > 0, \end{aligned} \quad (12.2)$$

where the following notation is introduced:

$$k^2 = \frac{2mE}{\hbar^2}, \quad k'^2 = \frac{2m}{\hbar^2}(E - U_0). \quad (12.3)$$

The solutions of these equations are respectively written in the form

$$\begin{aligned} \psi(x) &= A_1 e^{ikx} + B_1 e^{-ikx}, & x < 0, \\ \psi(x) &= A_2 e^{ik'x} + B_2 e^{-ik'x}, & x > 0. \end{aligned} \quad (12.4)$$

In these formulae a term of the form e^{ikx} represents a plane wave propagating in the positive direction of the x -axis, and e^{-ikx} represents a plane wave propagating in the opposite direction. The amplitudes A_1, B_1, A_2 and B_2 are integration constants. We define a flux of particles incident on the barrier. Let j_0 be the incident particle flux density. Then, according to (7.6),

$$j_0 = \frac{\hbar k}{m} |A_1|^2.$$

For simplicity we choose the flux such that we can set $A_1 = 1$.

In order to define the other constants we consider the behaviour of the wave function at the boundary of regions I and II at the point $x = 0$. By virtue of the general conditions imposed upon the wave function and its derivative (see §6), they must remain continuous even at the point of discontinuity of the potential energy. Hence for $x = 0$ the following equalities must hold:

$$\psi(+0) = \psi(-0), \quad (12.5)$$

$$\psi'(+0) = \psi'(-0). \quad (12.6)$$

From relations (12.5) and (12.6) one can also define the two integration constants A_2 and B_1 . As to the constant B_2 , we must set $B_2 = 0$. As a matter of fact, we only define the particle flux propagating in the positive direction of the x -axis. For $E > U_0$ (i.e. for real k') the term of the wave function proportional to $e^{-ik'x}$ represents a plane wave propagating in the opposite direction. The reflected wave propagates in region I in the negative direction of the x -axis. Obviously, the reflected wave is not present in region II and, consequently, the wave propagating from the right to the left is absent in this region. Hence we have to set the amplitude B_2 of this wave equal to zero. But if $E < U_0$ (k' is a purely imaginary quantity), then the function $e^{-ik'x}$ increases exponentially as $x \rightarrow +\infty$, which contradicts the condition of finiteness of the wave function. By virtue of this the coefficient B_2 must also be equal to zero for imaginary values of k' , i.e. for $E < U_0$.

Let us consider the case where the total energy of the particle is larger than the potential barrier height, $E > U_0$, in more detail.

Taking into account (12.4), from relation (12.5) and (12.6) we have

$$1 + B_1 = A_2, \quad k(1 - B_1) = k'A_2.$$

From these equations we find the amplitudes A_2 and B_1 :

$$B_1 = \frac{k - k'}{k + k'}, \quad A_2 = \frac{2k}{k + k'}. \quad (12.7)$$

We see that B_1 , the amplitude of the reflected wave, is different from zero, although $E > U_0$. This fact is due to the wave properties of particles. The wave is in part reflected, and in part passes into region II. The ratio of the flux density of reflected particles j_R to the flux density of incident particles j_0 will be called the reflection coefficient R . Correspondingly, the ratio of the flux density of transmitted particles j_D to the flux density of incident particles will be called the transmission coefficient D . Taking into account (12.4), we find

$$j_R = \frac{\hbar k}{m} |B_1|^2, \quad j_D = \frac{\hbar k'}{m} |A_2|^2.$$

Since $j_0 = \hbar k/m$, we obtain

$$R = \left(\frac{k - k'}{k + k'} \right)^2, \quad D = \frac{4kk'}{(k + k')^2}. \quad (12.8)$$

We see that the following relation is automatically fulfilled:

$$R + D = 1. \quad (12.9)$$

This relation expresses the law of conservation of the number of particles.

We note that the expressions (12.8) turn out to be symmetric with respect to k and k' , i.e. for particles of a given energy E the reflection coefficient (as well as the transmission coefficient) turns out to be independent of the direction of motion of the particles. Particles moving from the left to the right, i.e. against the action of the force at the point $x = 0$, have the same probability of being reflected at this point as particles of the same energy moving from the right to the left, in the direction of the action of the force at the point $x = 0$. This fact is also due to the wave character of the process and has a corresponding optical analogy.

Let us now consider the case $E < U_0$. Then k' is a purely imaginary quanti-

ty which is conveniently written in the form $k' = i\kappa$, where

$$\kappa = \frac{1}{\hbar} [2m(U_0 - E)]^{\frac{1}{2}}. \quad (12.10)$$

The amplitude of the reflected wave B_1 turns out to be a complex quantity, and the reflection coefficient R is equal to

$$R = |B_1|^2 = \left| \frac{k - i\kappa}{k + i\kappa} \right|^2 = 1, \quad D = 1 - R = 0. \quad (12.11)$$

The reflected wave is written in the form

$$\psi_R = \frac{k - i\kappa}{k + i\kappa} e^{-ikx} = e^{-i(kx + \delta)}, \quad (12.12)$$

i.e. the reflection leads to a shift of the phase of the wave. From (12.12) it follows that this shift is equal to

$$\delta = \arctan \frac{2k\kappa}{k^2 - \kappa^2}. \quad (12.13)$$

Although the reflection is complete, nevertheless the wave function in region II is different from zero and has the form

$$\psi(x) = A_2 e^{-\kappa x} = \frac{2k}{k + i\kappa} e^{-\kappa x} \quad (x > 0). \quad (12.14)$$

Correspondingly the probability density for finding the particle at point x in the region $x > 0$ is equal to

$$|\psi(x)|^2 = \frac{4k^2}{k^2 + \kappa^2} e^{-2\kappa x}. \quad (12.15)$$

We see that the behaviour of quantum particles differs essentially from that of classical particles. For a particle moving according to the laws of classical mechanics the region $x > 0$ for $E < U_0$ was forbidden. On the contrary, a particle moving according to the laws of quantum mechanics can, with a certain probability, penetrate this region. The penetration of particles into the region of forbidden energies represents a specific quantum effect which is called the tunnel effect. As is seen from the formula, the effective depth of penetration into region II, i.e. the distance δx from the boundary of region II at which the probability of finding the particle is still considerably different from zero, is of the order of magnitude of $\delta x \sim \kappa^{-1}$. For $x \gg \delta x$ the probability density (12.15) turns out to be exponentially small.

Let us estimate the effective depth of penetration for an electron, assum-

ing that $U_0 - E \sim 1 \text{ eV} = 1.6 \times 10^{-12} \text{ erg}$. For δx we evidently have

$$\delta x \sim \frac{\hbar}{[2m(U_0 - E)]^{\frac{1}{2}}} \sim \frac{10^{-27}}{[2 \times 10^{-27} \times 1.6 \times 10^{-12}]^{\frac{1}{2}}} \sim 10^{-8} \text{ cm}.$$

This estimate shows that the effect can be significant only in the realm of microscopic dimensions. Thus, as was to be expected, the tunnel effect cannot be observed in the motion of macroscopic bodies for which the laws of classical mechanics are valid.

In order to actually observe the particle in region II, we have to localize it there in a certain small interval $\Delta x \lesssim \delta x$. By localizing the particle we change its state (its energy), since by virtue of the uncertainty relation $\Delta p \gtrsim \hbar/\Delta x \gtrsim \hbar\kappa$. The particle which we shall detect somewhere in region II will no longer possess the initial energy E . The uncertainty in the momentum is related to the uncertainty in the kinetic energy of the particle:

$$\Delta T \gtrsim \frac{\Delta p^2}{2m} \gtrsim \frac{\hbar^2 \kappa^2}{2m}.$$

Substituting here the expression for κ (12.10), we obtain

$$\Delta T \gtrsim U_0 - E.$$

Thus the uncertainty in the energy of a particle localized in the region behind the barrier is larger than the energy that it lacks to reach the barrier height.

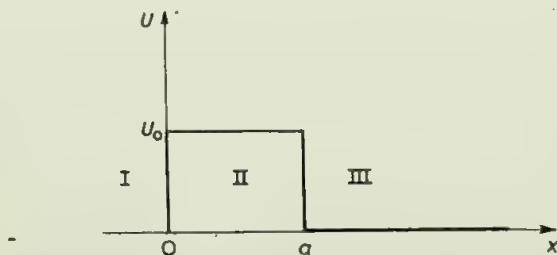


Fig. V.7

Let us consider briefly a barrier of finite extent, as shown in fig. V.7. Particles are incident on the barrier, moving in the positive direction of the x -axis. We can immediately write the wave function for the three different

regions:

$$\psi(x) = e^{ikx} + B_1 e^{-ikx}, \quad x < 0, \quad U = 0, \quad \text{I} \quad (12.16)$$

$$\psi(x) = A_2 e^{ik'x} + B_2 e^{-ik'x}, \quad 0 \leq x \leq a, \quad U = U_0 \quad \text{II} \quad (12.17)$$

$$\psi(x) = A_3 e^{ikx}, \quad x > a, \quad U = 0. \quad \text{III} \quad (12.18)$$

In this case (as in the case of a barrier of infinite extent) we again set the amplitude of the incident wave equal to unity. Since in region III there is no reflected wave, only the wave propagating in the positive direction of the x -axis is taken into account. We write the conditions of continuity of the wave function and of its first derivative at the boundaries of the regions, analogous to (12.5) and (12.6):

$$\begin{aligned} \psi(+0) &= \psi(-0), & \psi'(+0) &= \psi'(-0), \\ \psi(a+0) &= \psi(a-0), & \psi'(a+0) &= \psi'(a-0). \end{aligned} \quad (12.19)$$

Substituting (12.16), (12.17) and (12.18) into these relations, we obtain a system of equations with respect to B_1, A_2, B_2 and A_3

$$\begin{aligned} 1 + B_1 &= A_2 + B_2, & A_2 e^{ik'a} + B_2 e^{-ik'a} &= A_3 e^{ika}, \\ k(1 - B_1) &= k'(A_2 - B_2), & A_2 e^{ik'a} - B_2 e^{-ik'a} &= \frac{k}{k'} A_3 e^{ika}. \end{aligned} \quad (12.20)$$

Let us consider immediately the most interesting case, $E < U_0$. If the motion of particles proceeded according to the laws of classical mechanics, then the barrier would be completely impenetrable for them, and at the point $x = 0$ the particles would undergo total reflection from the barrier. The situation is different in the case of microparticles, whose motion is described by the laws of quantum mechanics. Solving the system (12.20) with respect to A_3 and taking into account that $k' = i\kappa$, where κ is defined by (12.10), we obtain

$$A_3 = \frac{4i\kappa k e^{-ika}}{(k + i\kappa)^2 e^{\kappa a} - (k - i\kappa)^2 e^{-\kappa a}}. \quad (12.21)$$

The amplitude of the plane wave turns out to be different from zero in the region behind the barrier, although the energy of the particle is smaller than the barrier height $E < U_0$. This means that a microparticle can, with a certain probability, pass through the potential barrier by means of the tunnel effect.

The tunnel transmission of particles, which earlier seemed to be paradoxical, is at present not only observed experimentally, but plays a fundamental

role in a number of fields of physics, in particular in nuclear physics. It suffices to say that tunnel passage through a barrier is associated with the α -decay of radioactive nuclei and with the phenomenon of spontaneous fission of uranium nuclei. The tunnel effect is also related to the phenomenon of emission of electrons from cold metals in a strong electric field and to a number of other processes.

Let us define the coefficient of transmission D of microparticles through a barrier

$$D = \frac{j_D}{j_0} = |A_3|^2 = \frac{4k^2\kappa^2}{(k^2 + \kappa^2)^2 \sinh^2 \kappa a + 4k^2\kappa^2}. \quad (12.22)$$

If $\kappa a \gg 1$, then $\sinh \kappa a \approx \frac{1}{2}e^{\kappa a}$ and the expression (12.22) is simplified:

$$D \approx \frac{16k^2\kappa^2}{(k^2 + \kappa^2)^2} e^{-2\kappa a}. \quad (12.23)$$

The basic dependence of the transmission coefficient on the width and height of the barrier is defined by the exponential function $e^{-2\kappa a}$. Denoting the factor multiplying this by D_0 , we have

$$D = D_0 \exp \left[-\frac{2}{\hbar} [2m(U_0 - E)]^{\frac{1}{2}} a \right]. \quad (12.24)$$

We see that the probability of passing through a barrier is not too low if

$$\frac{2}{\hbar} [2m(U_0 - E)]^{\frac{1}{2}} a \lesssim 1. \quad (12.25)$$

The condition (12.25) can evidently be fulfilled only in the realm of microphenomena. If we substitute into (12.25) values on a nuclear scale $a \sim 10^{-13}$ cm, $m \sim 10^{-24}$ g (nucleon mass), $U_0 - E \sim 10$ MeV (10^{-5} erg), then carrying out the estimation we find that $D \sim e^{-1}$. Thus the particle can with considerable probability pass through a barrier whose height exceeds its energy by 5–10 MeV. A completely different result is obtained for the same particle and the same barrier height, if the spatial extent of the barrier amounts to $a \sim 1$ cm. Then $D \sim 10^{-13}$. This means that in the realm of macroscopic phenomena the effect of tunnel transmission is virtually absent. The probability of tunnel passage through a barrier of arbitrary form will be considered in §42.

§13. One-dimensional motion*

In this chapter we have considered a number of simple problems of quantum mechanics regarding one-dimensional motion. The three-dimensional problems considered also reduce to one-dimensional problems, because the Schrödinger equation with the potential $U(x, y, z) = U_1(x) + U_2(y) + U_3(z)$ reduces to one-dimensional equations with the potentials U_1 , U_2 and U_3 respectively. The results obtained allow one to draw some general conclusions on the properties of the one-dimensional motion of particles.

For one-dimensional motion the energy levels can belong to a discrete spectrum (see §8–11) as well as to a continuous spectrum (§12). As we have seen, there correspond to the states of the discrete spectrum quadratically integrable wave functions, i.e. wave functions for which the normalization condition can be written in the form $\int |\psi_n(x)|^2 dx = 1$. This condition indicates that the motion is finite, i.e. that the probability of observing the particle at arbitrarily large distances is negligible.

On the contrary, if the particle can go off to arbitrarily large distances, i.e. if the motion is infinite, its wave function is not quadratically integrable. It can be shown that in this case the energy has a complex spectrum. Suppose that the potential energy $U(x)$ of the particle somehow varies from the value $U(\infty)$ for $x \rightarrow \infty$, which we shall choose as the zero energy, to the value $U(-\infty) = U_0$ for $x \rightarrow -\infty$. We shall assume for definiteness that U_0 is positive, $U_0 > 0$. The function $U(x)$ may vary quite arbitrarily. We suppose only that it has a minimum $U_{\min} < 0$. Then for an energy E such that $U_{\min} < E < 0$ the particle cannot go off to infinity. For these energy values the motion is finite, and the spectrum is discrete. The energy levels of a discrete spectrum are not degenerate. This statement is easily proved by assuming the contrary. Indeed, if it is assumed that ψ_1 and ψ_2 are two solutions of the Schrödinger equation corresponding to one and the same energy value E , then they satisfy the relation

$$\frac{1}{\psi_1} \frac{d^2 \psi_1}{dx^2} = \frac{2m}{\hbar^2} (U - E) = \frac{1}{\psi_2} \frac{d^2 \psi_2}{dx^2},$$

i.e.

$$\frac{1}{\psi_1} \frac{d^2 \psi_1}{dx^2} = \frac{1}{\psi_2} \frac{d^2 \psi_2}{dx^2}.$$

* For a detailed treatment see the book of L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

Integrating this relation with respect to x , we obtain

$$\psi_2 \frac{d\psi_1}{dx} - \psi_1 \frac{d\psi_2}{dx} = \text{const.} \quad (13.1)$$

But at infinity $\psi_1 = \psi_2 = 0$. Hence the constant on the right-hand side of relation (13.1) is equal to zero and, consequently,

$$\psi_2 \frac{d\psi_1}{dx} = \psi_1 \frac{d\psi_2}{dx}.$$

Integrating once more with respect to x , we obtain $\psi_2 = \text{const.} \psi_1$. This means that the two functions describe one and the same state, i.e. degeneracy is absent.

In the range $0 < E < U_0$ the particle can move arbitrarily distant in the direction of positive x . Therefore the motion is infinite, and the energy spectrum is continuous. The wave functions in this case are also not degenerate. In fact, the preceding proof is again valid, since the wave functions reduce to zero when $x \rightarrow -\infty$. The asymptotic expressions for the wave functions for $x \rightarrow \pm \infty$ are easily obtained from the Schrödinger equation

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} [E - U(x)] \psi = 0$$

if one substitutes here $U = 0$ for $x \rightarrow \infty$ and $U = U_0$ for $x \rightarrow -\infty$. Correspondingly we obtain

$$\psi(x) = A \sin(kx + \alpha) \quad \text{for } x \rightarrow \infty \quad (13.2)$$

$$\psi(x) = B e^{\kappa x} \quad \text{for } x \rightarrow -\infty \quad (13.3)$$

where

$$k = \frac{1}{\hbar} (2mE)^{\frac{1}{2}} \quad \text{and} \quad \kappa = \frac{1}{\hbar} [2m(U_0 - E)]^{\frac{1}{2}},$$

i.e. the solution has the form of a standing plane wave when $x \rightarrow \infty$, and is exponentially damped when $x \rightarrow -\infty$.

In the energy range $E > U_0$ the motion is infinite in both directions. The energy spectrum is continuous. Since the Schrödinger equation is of second order, it has two linearly independent solutions. In this energy range both solutions satisfy the necessary requirements. Therefore the energy levels are two-fold degenerate. The asymptotic expression for the wave function is of the form

$$\psi = A_1 e^{ikx} + A_2 e^{-ikx}, \quad (13.4)$$

where one term corresponds to the particle moving in the positive direction of the x -axis, and the other corresponds to the particle moving in the negative direction of the x -axis.

We now assume that the field increases indefinitely, i.e. that $|U(x)| \rightarrow \infty$ as $x \rightarrow \pm \infty$. As a simple and at the same time important example we shall consider the problem of the motion of a particle in a uniform external field

$$U(x) = -fx.$$

We choose the x -axis to be in the direction of the field; f denotes the force acting on the particle: $f = -dU/dx$. The potential energy is measured from its value at $x = 0$, hence $U(0) = 0$. The Schrödinger equation for the motion in such a field is of the form

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E + fx)\psi = 0. \quad (13.5)$$

We now introduce in place of x the new variable

$$\eta = \left(\frac{2mf}{\hbar^2}\right)^{\frac{1}{3}} \left(x + \frac{E}{f}\right).$$

Correspondingly eq. (13.5) will have the form

$$\frac{d^2\psi}{d\eta^2} + \eta\psi = 0. \quad (13.6)$$

The solution of eq. (13.6) can be expressed in terms of Bessel functions, but the expression of the solution of (13.6) in terms of the so-called Airy function is more convenient. Namely, the solution of eq. (13.6), finite for all values of η , has the form

$$\psi(\eta) = C\Phi(-\eta), \quad (13.7)$$

where $\Phi(\eta)$ denotes the Airy function,

$$\Phi(\eta) = \pi^{-\frac{1}{2}} \int_0^{\infty} \cos\left(\frac{1}{3}u^3 + u\eta\right) du, \quad (13.8)$$

and C is the normalization constant.

Thus the Schrödinger equation (13.5) has a solution satisfying the necessary requirements for any energy value E . Consequently, for motion in a uniform field the energy spectrum of the particle is continuous, which corres-

ponds to infinite motion. In the given case $U \rightarrow -\infty$ as $x \rightarrow \infty$, i.e. the motion in the positive direction of the x -axis is infinite.

The wave function (13.7) has a rather simple form for $\eta \rightarrow \pm \infty$. Making use of the known asymptotic expressions for the Airy function, we have

$$\psi(\eta) = \frac{C}{2|\eta|^{\frac{1}{4}}} \exp\left(-\frac{2}{3}|\eta|^{\frac{3}{2}}\right) \quad \text{for } \eta \rightarrow -\infty \quad (13.9)$$

and

$$\psi(\eta) = \frac{C}{\eta^{\frac{1}{4}}} \sin\left(\frac{2}{3}\eta^{\frac{3}{2}} + \frac{1}{4}\pi\right) \quad \text{for } \eta \rightarrow \infty. \quad (13.10)$$

The constant C is defined by the normalization condition. Of course, the integral of the square of the modulus of the wave function (13.7) over all space diverges, which corresponds to infinite motion. Normalization rules for such cases will be discussed in § 18.

§ 14. The Schrödinger equation for a system of particles

In the preceding sections we have considered the laws of motion of a single particle in an external field. But this greatly restricted the range of problems we could consider. As a matter of fact, even the simplest system, the hydrogen atom, represents, strictly speaking, a system of two particles. This holds all the more for systems such as many-electron atoms, molecules, atomic nuclei, matter in the solid state and so on. Generalizing the results obtained in § 6, we formulate the fundamental equation of quantum mechanics: the Schrödinger equation for a system of N particles. It has the form

$$i\hbar \frac{\partial \psi}{\partial t} = \sum_{i=1}^N \left(-\frac{\hbar^2}{2m_i} \right) \nabla_i^2 \psi + \sum_{i=1}^N U_i(\mathbf{r}_i) \psi + U_{\text{int}}(\mathbf{r}_1, \dots, \mathbf{r}_N) \psi. \quad (14.1)$$

Here the Laplacian ∇_i^2

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$$

acts on the coordinates of the i th particle. $U_i(\mathbf{r}_i)$ is the potential energy of the i th particle in the external field, U_{int} is the potential energy of the interaction of the particles with each other, and m_i is the mass of the i th particle. The summation is carried out over all particles of the system. The wave function

describing a system of particles, in accordance with §2, depends on the coordinates of all the particles and the time, $\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, t)$.

The Schrödinger equation for stationary states has the form

$$\sum_{i=1}^N \left(-\frac{\hbar^2}{2m_i} \right) \nabla_i^2 \psi + \sum_{i=1}^N U_i(\mathbf{r}_i) \psi + U_{\text{int}}(\mathbf{r}_1, \dots, \mathbf{r}_N) \psi = E \psi. \quad (14.2)$$

As the simplest example of the integration of eq. (14.2) let us consider a system of particles which do not interact with each other, i.e. let us assume that the energy of interaction is equal to zero, $U_{\text{int}} = 0$. In this case the Schrödinger equation can be rewritten in the form

$$\sum_{i=1}^N \left(-\frac{\hbar^2}{2m_i} \nabla_i^2 + U_i(\mathbf{r}_i) \right) \psi = E \psi. \quad (14.3)$$

where the terms in each bracket depend only on the coordinates of the corresponding particle. We seek a wave function ψ in the form of a product of functions which depend on the coordinate of individual particles.

$$\psi = \psi_1(\mathbf{r}_1) \psi_2(\mathbf{r}_2) \dots \psi_N(\mathbf{r}_N). \quad (14.4)$$

On substituting into the Schrödinger equation, we obtain

$$\begin{aligned} \sum_{i=1}^N \psi_1(\mathbf{r}_1) \dots \psi_{i-1}(\mathbf{r}_{i-1}) \psi_{i+1}(\mathbf{r}_{i+1}) \dots \\ \dots \psi_N(\mathbf{r}_N) \left(-\frac{\hbar^2}{2m_i} \nabla_i^2 + U_i(\mathbf{r}_i) \right) \psi_i(\mathbf{r}_i) = E \psi. \end{aligned}$$

Dividing the right and left sides of the equation by ψ , we find

$$\sum_{i=1}^N \frac{1}{\psi_i(\mathbf{r}_i)} \left(-\frac{\hbar^2}{2m_i} \nabla_i^2 + U_i(\mathbf{r}_i) \right) \psi_i(\mathbf{r}_i) = E.$$

There is a constant quantity on the right-hand side of the equation. The left-hand side is made up of a sum of terms each of which is a function of its independent variable. In order that the equation hold for all values of the independent variables, the following conditions must be fulfilled:

$$-\frac{\hbar^2}{2m_i} \nabla_i^2 \psi_i(\mathbf{r}_i) + U_i(\mathbf{r}_i) \psi_i(\mathbf{r}_i) = E_i \psi_i(\mathbf{r}_i), \quad \sum_{i=1}^N E_i = E,$$

where the E_i are constants, which, as is easily seen, represent the energies of individual particles.

Thus, if the left-hand side of the Schrödinger equation can be written in the form of the sum (14.3), then the wave function of the system resolves into a product of wave functions, while the energy of the system appears as the sum of the energies of the individual particles.

These results have a simple physical meaning. We have assumed that the energy of interaction between the particles is equal to zero. It is therefore natural that the total energy of the entire system is made up of the sum of the energies of the individual particles, since the motion of each of the particles is independent of the motion of the other particles. The probability of observing a given set of coordinates of the particles is written in the form

$$dW(\mathbf{r}_1, \dots, \mathbf{r}_N) = |\psi_1(\mathbf{r}_1)|^2 |\psi_2(\mathbf{r}_2)|^2 \dots |\psi_N(\mathbf{r}_N)|^2 dV_1 \dots dV_N.$$

The above result is in complete agreement with the theorem of multiplication of probabilities of independent events.

Further, let us consider in more detail a system of two particles with masses m_1 and m_2 . We assume that the potential energy of interaction depends only on the distance between the particles, and that there is no external field. In this case the Schrödinger equation for stationary states has the form

$$\begin{aligned} -\frac{\hbar^2}{2m_1} \nabla_1^2 \psi(\mathbf{r}_1, \mathbf{r}_2) - \frac{\hbar^2}{2m_2} \nabla_2^2 \psi(\mathbf{r}_1, \mathbf{r}_2) + \\ + U(|\mathbf{r}_1 - \mathbf{r}_2|) \psi(\mathbf{r}_1, \mathbf{r}_2) = E \psi(\mathbf{r}_1, \mathbf{r}_2). \end{aligned} \quad (14.5)$$

We transform this equation by introducing new coordinates \mathbf{R} and \mathbf{r} which are defined by the relation

$$\mathbf{R} = \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{m_1 + m_2}, \quad \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2. \quad (14.6)$$

We note that the new variables are completely analogous to the coordinates of the centre of mass and those of relative motion in classical mechanics. As a result of somewhat lengthy but not complicated transformations the Schrödinger equation takes the form

$$-\frac{\hbar^2}{2M} \nabla_{\mathbf{R}}^2 \psi - \frac{\hbar^2}{2\mu} \nabla_{\mathbf{r}}^2 \psi + U(r) \psi = E \psi. \quad (14.7)$$

Here M and μ are the total and reduced masses of the system

$$M = m_1 + m_2, \quad \mu = \frac{m_1 m_2}{m_1 + m_2}. \quad (14.8)$$

We see that the left-hand side of the Schrödinger equation is resolved into a sum of two terms and has a form similar to eq. (14.3). In this case the Schrödinger equation can be written in the form

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \varphi(\mathbf{R}) \psi_0(\mathbf{r}). \quad (14.9)$$

Substituting (14.9) into (14.7) and repeating the transformations which were carried out before, we obtain

$$-\frac{\hbar^2}{2M} \nabla_{\mathbf{R}}^2 \varphi = E_{\mathbf{R}} \varphi, \quad (14.10)$$

$$-\frac{\hbar^2}{2\mu} \nabla_{\mathbf{r}}^2 \psi_0 + U(\mathbf{r}) \psi_0 = E_{\mathbf{r}} \psi_0, \quad (14.11)$$

$$E_{\mathbf{R}} + E_{\mathbf{r}} = E. \quad (14.12)$$

Equation (14.10) is the Schrödinger equation for a free particle with mass M . Its solution is the function

$$\varphi(\mathbf{R}) = A e^{(i/\hbar)(\mathbf{P} \cdot \mathbf{R})}. \quad (14.13)$$

The quantity $E_{\mathbf{R}} = |\mathbf{P}|^2/2M$ represents the kinetic energy of motion of the system as a whole.

Thus corresponding to (14.9) and (14.13) the solution of the Schrödinger equation can be written in the form

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = A e^{(i/\hbar)(\mathbf{P} \cdot \mathbf{R})} \psi_0(\mathbf{r}). \quad (14.14)$$

From the formulae obtained it is seen that the centre of mass of the system moves in space as a free particle, while the relative motion of the particles proceeds independently of the motion of the centre of mass and is described by the function ψ_0 satisfying eq. (14.11). The total energy of the system is made up of the energies of the relative motion and the motion of the centre of mass. Hence we see that in quantum mechanics, as in classical physics, the problem of the motion of two particles whose potential energy of interaction U depends only on the distance between them, $U(|\mathbf{r}_1 - \mathbf{r}_2|)$, reduces to the problem of the motion of a single particle with reduced mass μ in an external field U .

The Mathematical Apparatus of Quantum Mechanics

§ 15. Linear operators

We have seen above that in solving the Schrödinger equation one can find the wave functions and the energy of a system. The latter in certain cases (e.g. a particle in a potential well) has a discrete sequence of values, while in other cases (e.g. a free particle; a particle passing through a barrier) it has a continuous sequence of values.

Knowing the wave function ψ , we could obtain the probability of finding the particle at a given point of space, as well as the mean values of the quantities depending on the coordinates. Then, as was shown in §4, the coordinates and the corresponding momentum components of the particle have no definite simultaneous values. However, the mathematical apparatus we used was inadequate to solve a number of important problems. As examples we put these questions: Which quantities cannot simultaneously have definite values? How are the mean values of the quantities which are not functions of the coordinates to be found? What characteristics of a quantum-mechanical system must be given in order that its state is completely defined?

The peculiarity of the problems of quantum mechanics required the development and application of a special mathematical technique.

The mathematical technique of quantum mechanics must correspond to

the physical statement of its problems. It turned out that the corresponding mathematical technique — the theory of linear operators — had already been worked out in mathematics. We shall first consider the basic concepts of this theory, and in what follows we shall show how the theory of linear operators can be associated with the problems of quantum mechanics.

We shall understand by an operator a rule by which a function $\varphi(x_1, x_2, x_3, \dots)$ of the variables x_1, x_2, x_3, \dots is related to another function $\chi(x_1, x_2, x_3, \dots)$ of the same variables.

In what follows we shall denote operators by means of letters with the sign $\hat{}$, for example \hat{F} . By means of the symbol \hat{F} the rule of the transition from the function φ to the function χ can be written in the form

$$\chi = \hat{F}\varphi. \quad (15.1)$$

Let us consider some simple operators.

The operator \hat{F} can, for example, denote differentiation with respect to any variable

$$\chi(x_1, x_2, \dots) = \frac{\partial}{\partial x_i} \varphi(x_1, x_2, \dots).$$

This operator is written symbolically as follows:

$$\hat{F} = \frac{\partial}{\partial x_i}.$$

The differential operators will be encountered particularly often. The operator \hat{F} can also denote multiplication by any quantity, raising to a power, and so on.

We define the operator of the independent variable x as the multiplication by this variable: $\chi = x\varphi$.

An integral relation between the functions φ and χ can also be represented in operator form:

$$\chi(x) = \int K(x, \xi) \varphi(\xi) d\xi = \hat{F}\varphi. \quad (15.2)$$

The function $K(x, \xi)$ is called the kernel of the integral operator \hat{F} .

We note that earlier we have also used differential operators: the operator $\nabla = i(\partial/\partial x) + j(\partial/\partial y) + k(\partial/\partial z)$, the Laplacian $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ and others.

We now define a linear operator \hat{F} as an operator for which the equalities

$$\hat{F}(\varphi_1 + \varphi_2) = \hat{F}\varphi_1 + \hat{F}\varphi_2, \quad (15.3)$$

$$\hat{F}C\varphi = C\hat{F}\varphi \quad (15.4)$$

are fulfilled, where C is an arbitrary constant. Hence it follows that

$$\hat{F}(C_1\varphi_1 + C_2\varphi_2) = C_1\hat{F}\varphi_1 + C_2\hat{F}\varphi_2, \quad (15.5)$$

where C_1, C_2 are arbitrary constants. It is obvious that the operators mentioned above are linear operators.

For reasons which will be clear from what follows we shall, in quantum mechanics, deal only with linear operators.

By combining two given operators \hat{F} and \hat{R} one can define their sum and product. We shall understand the sum of the operators \hat{F} and \hat{R} to be the operator \hat{G} defined by the relation

$$\hat{G} = \hat{F} + \hat{R}, \quad \hat{G}\varphi = \hat{F}\varphi + \hat{R}\varphi. \quad (15.6)$$

We shall understand the product of two operators \hat{F} and \hat{R} to be the operator $\hat{L} = \hat{F}\hat{R}$ consisting of the consecutive application of the operators \hat{R} and \hat{F} ,

$$\hat{L}\varphi = \hat{F}(\hat{R}\varphi). \quad (15.7)$$

If one first applies the operator \hat{F} and then the operator \hat{R} , their product will be the operator $\hat{L}' = \hat{R}\hat{F}$,

$$\hat{L}'\varphi = \hat{R}(\hat{F}\varphi). \quad (15.8)$$

We note that the operators \hat{L} and \hat{L}' , generally speaking, are not identical to each other, i.e. the product of the operators depends critically on the order of the factors. Corresponding to this, the algebra of operators is an algebra of non-commuting quantities. Two operators are said to commute with each other if the product of the operators does not depend on the order of the factors; otherwise the operators are said to be non-commuting. As an example let us find the product of the operator of differentiation with respect to x and the operator of multiplication by x for both orders of the factors, i.e. let us assume that $\hat{F} = x$, $\hat{R} = d/dx$. In correspondence with (15.7) the operator of the product \hat{L} will be $\hat{L} = x(d/dx)$. We now find the operator $\hat{L}' = \hat{R}\hat{F}$

$$\hat{L}'\psi = \frac{\partial}{\partial x}(x\psi) = \left(1 + x \frac{\partial}{\partial x}\right)\psi.$$

We see that in this case the operator \hat{L}' is equal to

$$\hat{L}' = 1 + x \frac{\partial}{\partial x}$$

and is not identical to the operator \hat{L} . Thus the operators \hat{L} and \hat{L}' do not commute. Making use of the expressions obtained for \hat{L} and \hat{L}' , we can write

that

$$\frac{\partial}{\partial x} x - x \frac{\partial}{\partial x} = 1.$$

It is natural to call the right-hand side of this operator relation the unit operator. If we took the operator of multiplication by any other independent variable, say y , as the operator \hat{F} , then it would turn out that the operators $\partial/\partial x$ and y commute

$$y \frac{\partial}{\partial x} - \frac{\partial}{\partial x} y = 0. \quad (15.9)$$

For certain operators it turns out that the following relation holds:

$$\hat{F}\hat{R} = -\hat{R}\hat{F}. \quad (15.10)$$

In this case the operators \hat{F} and \hat{R} are said to anticommute. We shall call the operator $\hat{F}\hat{R} - \hat{R}\hat{F}$ the commutator of the operators \hat{F} and \hat{R} , and shall denote it by brackets, i.e.

$$\hat{F}\hat{R} - \hat{R}\hat{F} = \{\hat{F}, \hat{R}\}. \quad (15.11)$$

An operator \hat{F} can be contrasted with the inverse operator \hat{F}^{-1} . The inverse operator is defined by the relations

$$\hat{F}^{-1}\hat{F}\psi = \psi, \quad \hat{F}\hat{F}^{-1}\psi = \psi$$

or

$$\hat{F}\hat{F}^{-1} = \hat{F}^{-1}\hat{F} = 1. \quad (15.12)$$

If \hat{F} is a differential operator, then the inverse operator \hat{F}^{-1} has the form of an integral operator. Indeed, suppose that the relation

$$\hat{F}\psi(x) = \varphi(x). \quad (15.13)$$

holds. Then, acting on the right-hand and left-hand sides of this equality with the operator \hat{F}^{-1} , we obtain

$$\psi(x) = \hat{F}^{-1}\varphi(x). \quad (15.14)$$

On the other hand, the relation (15.14) can be written in the form

$$\psi(x) = \int G(x, x')\varphi(x')dx', \quad (15.15)$$

where the function $G(x, x')$, called the Green's function of eq. (15.13),

satisfies the relation

$$\hat{F}G(x, x') = \delta(x - x'). \quad (15.16)$$

Indeed, if we act with the operator \hat{F} on the right-hand and left-hand sides of the equality (15.15), then under the condition (15.16) we again arrive at the relation (15.13). Comparing (15.14) and (15.15), we see that the Green's function $G(x, x')$ is the kernel of the integral operator \hat{F}^{-1} . The Green's function $G(x, x')$ is not determined unambiguously from eq. (15.16). For a single-valued determination it is necessary to give in addition certain conditions of the nature of boundary conditions.

The relation (15.13) can be considered as an equation relating the function $\psi(x)$ to a given function $\varphi(x)$, and whose solution is given by formula (15.15). It should only be borne in mind that in order to obtain a general solution we have to add the general solution $\psi_0(x)$ of the homogeneous equation $\hat{F}\psi_0(x) = 0$ to (15.15). We then have

$$\psi(x) = \psi_0(x) + \int G(x, x') \varphi(x') dx'. \quad (15.17)$$

We shall need the above relation in what follows.

§16. Eigenvalues and eigenfunctions of operators

Let us consider the operator relation

$$\hat{F}\psi = F\psi. \quad (16.1)$$

Relation (16.1) means that if the operator \hat{F} is applied to the function ψ , then one again obtains the function ψ multiplied by a certain constant F . It is obvious that for a given form of the operator \hat{F} the relation (16.1) cannot be satisfied by every function ψ . In other words, relation (16.1) is an equation. The form of the function ψ can be obtained by solving eq. (16.1). If the operator \hat{F} is a linear differential operator, then eq. (16.1) will be a differential equation. Since from the form of the equation it is immediately clear that $\psi = 0$ is its trivial solution, (16.1) represents a linear homogeneous differential equation. The study of such linear homogeneous equations is the most important problem of the theory of operators.

In what follows we shall not be interested in arbitrary operators \hat{F} and arbitrary functions ψ , but only in functions which satisfy certain definite conditions:

(1) the function ψ must exist over the entire range of the independent varia-

bles. For example, in the case of Cartesian coordinates, in the range $-\infty < x < \infty$, $-\infty < y < \infty$, $-\infty < z < \infty$;

(2) in the region of existence the function ψ must be finite and continuous, with the exception, in some cases, of singular points;

(3) the function ψ must be single-valued.

We shall call the set of conditions (1)–(3) the standard conditions. It turns out, generally speaking, that eq. (16.1) has solutions which differ from the trivial one, and which satisfy the standard conditions, not for all values of the parameter F but only for certain selected values of it. The selected values of F for which the non-trivial solutions of eq. (16.1) exist are called the eigenvalues of the operator \hat{F} , and the corresponding solutions of eq. (16.1) are called the eigenfunctions of the operator \hat{F} .

We shall first of all present the problems on eigenfunctions and eigenvalues with which we are already acquainted.

(1) In considering the problem of the motion of a particle in a potential well we have solved eq. (16.1) with the differential operator $\hat{F} = -d^2/dx^2$. The boundary conditions led to the eigenvalues (8.5) and to the eigenfunctions (8.7) of the operator \hat{F} .

(2) If, for the same form of the operator, we require no reduction of ψ to zero at the boundaries of the interval $(0, l)$, then the solutions of (8.2) will have the form

$$\psi = Ae^{ikx} + Be^{-ikx}.$$

If $k^2 > 0$, then for all values of x the function ψ is finite, so that the solution satisfies the standard conditions. For negative k^2

$$\psi = Ae^{-\kappa x} + Be^{\kappa x}, \quad \text{where } \kappa = ik,$$

there are no solutions which satisfy the standard conditions.

(3) In the problem of the oscillator we have considered the solution of eq. (16.1) for the operator (see (10.3))

$$\hat{F} = -\frac{d^2}{dx^2} + x^2.$$

The problem has the solution $F = 2E/\hbar\omega = 2n + 1$.

It is clear from the examples given that the whole set of eigenvalues of an operator, which we shall call its spectrum, can be discrete (example 1 and 3) as well as continuous (example 2). It can be proved that the eigenfunctions which correspond to the discrete spectrum of the eigenvalues are quadratically integrable, i.e. the integral $\int |\psi|^2 dV$ converges. The eigenfunctions corresponding to the continuous spectrum of the eigenvalues are not quadratically

integrable. If to each eigenvalue of the operator there belongs one and only one eigenfunction ψ , the spectrum is said to be non-degenerate. On the contrary, if to one eigenvalue F there correspond several, for example s , different eigenfunctions, then the given eigenvalue is said to be degenerate with degeneracy s .

The examples given above are important because they illuminate our interest in the theory of operators. The problem of finding the solutions of the Schrödinger equation is a particular case of the problem of the eigenfunctions of operators of a particular form.

Before passing over from this heuristic reasoning to the establishment of a more complete relationship between the concepts of quantum mechanics and the theory of linear operators, it is necessary in addition to consider important properties of a particular class of operators.

§ 17. Hermitian operators

The eigenvalues F in the operator equation (16.1) can, generally speaking, be complex. However, we shall be solely interested in equations which lead only to real eigenvalues. It turns out that there is a class of operators which can possess only real eigenvalues. Such operators are called Hermitian or self-adjoint. Each linear operator \hat{F} can be compared to a certain operator \hat{F}^\dagger which we shall call the adjoint operator or the Hermitian conjugate. The adjoint operator is defined by the condition

$$\int \psi_1^* \hat{F} \psi_2 dV = \int \psi_2 (\hat{F}^\dagger \psi_1)^* dV. \quad (17.1)$$

Here, as always, the asterisk denotes complex-conjugate quantities. The integration in (17.1) is carried out over the entire region of variation of the independent variables. We have denoted by dV a volume element of this region.

The functions ψ_1, ψ_2 must satisfy the necessary requirements for the convergence of the integrals in (17.1). Furthermore, they must satisfy certain boundary conditions which usually amount to the requirement that the functions ψ_1 and ψ_2 reduce to zero at infinity. But in other respects the functions ψ_1 and ψ_2 are rather arbitrary. If the operator \hat{F} coincides with its adjoint operator $\hat{F}^\dagger = \hat{F}$, then such an operator is said to be Hermitian or self-adjoint. In this case relation (17.1) has the form

$$\int \psi_1^* \hat{F} \psi_2 dV = \int \psi_2 \hat{F}^* \psi_1^* dV. \quad (17.2)$$

Here we have denoted by \hat{F}^* the operator defined by the relation $\hat{F}^* \psi^* = (\hat{F} \psi)^*$.

As an example let us find the adjoint operator of the differential operator $\hat{F} = d/dx$. Assuming that the functions ψ_1, ψ_2 reduce to zero at infinity and carrying out the integration by parts in (17.1), we obtain

$$\int_{-\infty}^{\infty} \psi_1^* \frac{d}{dx} \psi_2 dx = - \int_{-\infty}^{\infty} \psi_2 \frac{d}{dx} \psi_1^* dx.$$

Comparing with (17.1), we find the operator \hat{F}^\dagger : $\hat{F}^\dagger = -d/dx$. We see that the operator \hat{F}^\dagger in the given case does not coincide with the operator \hat{F} , i.e. the differential operator is not self-adjoint. If, however, the operator $i(d/dx)$ is taken as the operator \hat{F} , then it is easily seen that such an operator is Hermitian. Indeed, in this case we have, on integrating by parts

$$i \int_{-\infty}^{\infty} \psi_1^* \frac{d}{dx} \psi_2 dx = -i \int_{-\infty}^{\infty} \psi_2 \frac{d}{dx} \psi_1^* dx,$$

and relation (17.2) is now valid. Hence it follows that the operator $\hat{F} = i(d/dx)$ is Hermitian.

We also define the operator $\tilde{\hat{F}}$, which is called the transpose of the operator \hat{F}

$$\int \psi_1^* \hat{F} \psi_2 dV = \int \psi_2 \tilde{\hat{F}} \psi_1^* dV. \quad (17.3)$$

Comparing (17.3) with (17.1), we obtain $\hat{F}^\dagger = \tilde{\hat{F}}^*$.

Further we find the operator \hat{L}^\dagger adjoint of the operator \hat{L} which is the product of two operators $\hat{L} = \hat{F}\hat{R}$. From the definition (17.1) we have

$$\int \psi_1^* \hat{F}(\hat{R}\psi_2) dV = \int (\hat{R}\psi_2)(\hat{F}^\dagger \psi_1^*) dV.$$

We exchange the functions $(\hat{R}\psi_2)$ and $(\hat{F}^\dagger \psi_1^*)$. Then we get

$$\int \psi_1^* \hat{F}\hat{R}\psi_2 dV = \int (\hat{F}^\dagger \psi_1^*)^* \hat{R}\psi_2 dV.$$

Further, we again use the relation (17.1)

$$\int \psi_1^* \hat{F}\hat{R}\psi_2 dV = \int \psi_2 (\hat{R}^\dagger (\hat{F}^\dagger \psi_1^*))^* dV.$$

From this expression we obtain the operator $\hat{L}^\dagger = (\hat{F}\hat{R})^\dagger$

$$(\hat{F}\hat{R})^\dagger = \hat{R}^\dagger \hat{F}^\dagger. \quad (17.4)$$

We see that the adjoint operator of the product is equal to the product of the

adjoint operators taken, however, in the reverse order. Thus if the operators \hat{F} and \hat{R} are self-adjoint, i.e. $\hat{F}^\dagger = \hat{F}$, $\hat{R}^\dagger = \hat{R}$, then their product will be a self-adjoint operator only in the case where they commute. Indeed, under these conditions we have

$$(\hat{F}\hat{R})^\dagger = \hat{R}\hat{F} = \hat{F}\hat{R}. \quad (17.5)$$

Since each operator undoubtedly commutes with itself, it follows from (17.5) that if the operator \hat{F} is Hermitian, then so also will be the operator $\hat{F}^2 = \hat{F}\hat{F}$ as well as, in general, the operator $\hat{F}^n = \underbrace{\hat{F} \cdot \hat{F} \dots \hat{F}}_n$, where n is a positive integer.

We now pass on to the proof of the basic theorem on the reality of the eigenvalues of Hermitian operators. For this we once again write eq. (16.1), assuming for concreteness that the operator \hat{F} possesses a discrete spectrum $\hat{F}\psi_n = F_n\psi_n$. Multiplying the equation from the left by ψ_n^* and integrating, we obtain

$$F_n = \frac{\int \psi_n^* \hat{F} \psi_n dV}{\int |\psi_n|^2 dV}.$$

If the operator \hat{F} is Hermitian, then it is easily seen that the eigenvalues F_n , determined in (16.1), are real. Indeed, taking into account the Hermitian property (17.2), we find

$$F_n^* = \frac{\int \psi_n \hat{F}^* \psi_n^* dV}{\int |\psi_n|^2 dV} = F_n.$$

Thus we have proved that Hermitian (self-adjoint) operators have real eigenvalues only.

§18. The orthogonality and normalization of the eigenfunctions of Hermitian operators

The eigenfunctions of a linear Hermitian operator \hat{F} , which correspond to different eigenvalues F_n and F_m , are mutually orthogonal, i.e. satisfy the relation

$$\int \psi_m^* \psi_n dV = 0 \quad (\text{for } m \neq n). \quad (18.1)$$

The functions ψ_n and ψ_m^* satisfy eq. (16.1),

$$\hat{F}\psi_n = F_n\psi_n, \quad \hat{F}^*\psi_m^* = F_m\psi_m^*. \quad (18.2)$$

Since the operator F is Hermitian, we have:

$$\int \psi_m^* \hat{F} \psi_n dV = \int \psi_n \hat{F}^* \psi_m^* dV. \quad (18.3)$$

Making use of eqs. (18.2), we rewrite eq. (18.3) in the form

$$F_n \int \psi_m^* \psi_n dV = F_m \int \psi_m^* \psi_n dV.$$

Hence it follows that

$$(F_m - F_n) \int \psi_m^* \psi_n dV = 0. \quad (18.4)$$

Since by assumption $F_m \neq F_n$, then

$$\int \psi_m^* \psi_n dV = 0, \quad (18.5)$$

which proves our statement.

Because the eigenfunctions satisfy a homogeneous linear equation, they are determined to within an arbitrary constant.

Keeping in mind what is to follow, we shall normalize the eigenfunctions of a discrete spectrum by the condition

$$\int \psi_n^* \psi_n dV = 1. \quad (18.6)$$

The eigenfunctions satisfying relation (18.6) are said to be normalized to unity. We combine formulae (18.1) and (18.6) into one formula

$$\int \psi_m^* \psi_n dV = \delta_{nm}, \quad (18.7)$$

where δ_{nm} is the Kronecker symbol:

$$\delta_{nm} = \begin{cases} 1 & \text{if } n = m, \\ 0 & \text{if } n \neq m. \end{cases}$$

We now consider the case of degenerate states, where several eigenfunctions $\psi_{n1}, \psi_{n2}, \dots, \psi_{ns}$, belong to the same eigenvalue F_n .

One can take as the solution of eq. (18.2) corresponding to the eigenvalue F_n arbitrary linear combinations of these functions

$$\psi'_{nk} = \sum_{r=1}^s a_{kr} \psi_{nr}. \quad (18.8)$$

By appropriate choice of the coefficients a_{kr} one can obtain mutual orthogonality of the eigenfunctions ψ'_{nk} which belong to one and the same eigenvalue F_n . Imposing also the normalization condition, we obtain

$$\int \psi'_{nk}{}^* \psi'_{nl} dV = \delta_{kl}. \quad (18.9)$$

The condition (18.9) still does not completely determine the values of the coefficients a_{kr} . Indeed, if the functions ψ_{nk} are already orthogonal to each other and if we have carried out transformation (18.8), then the orthogonality will be preserved if

$$\sum_{r=1}^s a_{kr}^* a_{lr} = \delta_{kl}. \quad (18.10)$$

Thus there is still a certain arbitrariness in the choice of the coefficients a_{kl} .

Finally, we consider the wave functions of a continuous spectrum. For the wave functions of continuous spectrum $\psi_F(x)$ the condition of orthogonality is proved analogously to (18.3)–(18.5):

$$\int \psi_F^*(x) \psi_{F'}(x) dV = 0. \quad (18.11)$$

On the other hand, the condition of normalization can no longer be written in the form of (18.6), because the wave functions of the continuous spectrum are not quadratically integrable. For these functions the integral $\int |\psi_F|^2 dV$ diverges. This divergence is associated with the fact that the eigenfunctions of the continuous spectrum do not reduce to zero at infinity. The eigenfunctions of the continuous spectrum are conveniently normalized to the Dirac δ -function (see Vol. I, Appendix III), since the conditions of orthogonality and normalization can be expressed analogously to (18.7),

$$\int \psi_F^*(x) \psi_{F'}(x) dV = \delta(F - F'). \quad (18.12)$$

The normalization to the δ -function, of course, is not the only possible one. Later we shall encounter other methods of normalizing the eigenfunctions of the continuous spectrum (see, for example, § 26).

§ 19. Expansion in terms of eigenfunctions

In the preceding section we have proved that the system of eigenfunctions of an arbitrary linear self-adjoint operator is a system of orthogonal functions. It turns out that such a system of functions is complete. An arbitrary continuous function, determined in the same region of variation of the independent variables and satisfying a wide class of conditions, can be expanded in this set of eigenfunctions*.

We shall first give here the conditions for the completeness of a system of

* V.I. Smirnov, *A course of higher mathematics* (Pergamon Press, Oxford, 1964).

eigenfunctions for the case of an operator \hat{F} which possesses a discrete spectrum. We write the expansion of a function ψ in a series in terms of the eigenfunctions ψ_n , assuming the latter to be normalized to unity, in the form

$$\psi(x) = \sum_n c_n \psi_n(x). \quad (19.1)$$

The amplitudes c_n can be determined by making use of the orthogonality of the eigenfunctions. Multiplying (19.1) by $\psi_m^*(x)$ and integrating over all the regions of variation of the independent variables, we obtain

$$\int \psi_m^*(x) \psi(x) dV = \sum_n c_n \int \psi_m^*(x) \psi_n(x) dV.$$

Here we have changed the order of summation and integration. By virtue of the orthogonality of the eigenfunctions (18.7), of all terms of the sum on the right-hand side of the equation only the term with $n = m$ is different from zero. Consequently we have

$$c_m = \int \psi_m^*(x) \psi(x) dV. \quad (19.2)$$

Substituting this expression into (19.1) and again changing the order of summation and integration, we obtain

$$\psi(x) = \int \psi(x') \left(\sum_n \psi_n^*(x') \psi_n(x) \right) dV'. \quad (19.3)$$

For this expression to hold for an arbitrary continuous function $\psi(x)$ it is necessary that the equality

$$\sum_n \psi_n^*(x') \psi_n(x) = \delta(x - x') \quad (19.4)$$

be fulfilled. The relation (19.4) expresses the condition for the completeness of a system of eigenfunctions $\psi_n(x)$. If the operator \hat{F} possesses a continuous spectrum, then the expansion of the function $\psi(x)$ in terms of its eigenfunctions will be no longer a sum but an integral:

$$\psi(x) = \int c(F) \psi_F(x) dF. \quad (19.5)$$

The amplitudes $c(F)$ are found in the same way as in the case of a discrete spectrum. Multiplying the left-hand and right-hand sides of eq. (19.5) by the function $\psi_F^*(x)$ and integrating over the entire region of variation of the independent variables, we find

$$\int \psi_F^*(x) \psi(x) dV = \int c(F) dF \int \psi_F^*(x) \psi_F(x) dV.$$

Assuming that the eigenfunctions $\psi_F(x)$ are normalized to the δ -function,

we obtain finally

$$c(F) = \int \psi_F^*(x) \psi(x) dV. \quad (19.6)$$

We have already encountered in §3 a particular case of such an expansion (expansion in terms of plane waves). The condition for completeness in the case of a continuous spectrum is written analogously to (19.4)

$$\int \psi_F^*(x') \psi_F(x) dF = \delta(x - x'). \quad (19.7)$$

§20. Quantum-mechanical variables and operators

We can now turn to the discussion of the basic postulate of quantum mechanics which establishes the connection between real physical quantities which characterize the properties of quantum-mechanical systems and the mathematical apparatus of quantum mechanics.

In classical mechanics the state of a system is determined by the whole set of coordinates and momenta (or variables expressed in terms of the latter) involved in the equations of motion. All variables characterizing the state of a system are called mechanical variables. In quantum mechanics, variables which play an analogous role will be called quantum-mechanical variables. They are also often said to be physical or dynamical variables.

In the examples considered above certain properties of quantum systems have been elucidated. These are, in the first place:

- (1) the existence of an uncertainty relation between the values of canonically conjugate physical variables (such as, for example, the coordinate and momentum);
- (2) the existence of a discrete spectrum and a continuous spectrum of values of physical variables (for example, the energy of a quantum oscillator and of a free particle);
- (3) the existence of a superposition of quantum states (for example, the superposition of states of a free particle);
- (4) the continuous transition from the concepts of quantum mechanics to those of classical mechanics in passing to systems in which the Planck constant can be assumed to be an infinitesimal quantity, while quantum numbers can be assumed to be infinitely large (the correspondence principle).

The first and second of these properties just correspond to the properties of linear operators — their non-commutativity and the existence of a spectrum of eigenvalues. Hence it is natural to make the following basic assumption: 'To each quantum-mechanical variable F there corresponds a certain linear

Hermitian operator \hat{F} . The spectrum of eigenvalues of the operator \hat{F} represents the spectrum of the possible (measured) values of this variable'.

The eigenfunction $\psi_F(x)$ of the operator \hat{F} represents the wave function of the system in the state in which the variable represented by the operator \hat{F} has the given definite value F .

The requirement of hermiticity of the operator is connected, obviously, with the reality of the values of real physical quantities, whereas the requirement of linearity is associated with the principle of superposition. It is clear that this statement will assume a concrete meaning only after being supplemented by the indication of how the operator corresponding to a given quantum-mechanical quantity can be found. If such a recipe were known, then the postulate formulated would make it possible to determine the spectrum of the possible values of this quantity. The validity of the basic postulate can be established only by the agreement between the inferences of quantum mechanics and experiment.

For the determination of the form of the linear operators which correspond to definite quantum-mechanical variables — quantum-mechanical operators — it is necessary to make use of the correspondence principle. Namely, it is natural to assume that between quantum-mechanical operators describing the motion of particles in quantum mechanics there are the same relations as between their 'originals', the variables of classical mechanics. Thus, for example, the total energy operator \hat{H} is connected with the kinetic energy operator \hat{T} and the potential energy operator \hat{U} by the relation

$$\hat{H} = \hat{T} + \hat{U}. \quad (20.1)$$

In its turn the operator \hat{T} is equal to

$$\hat{T} = \frac{|\hat{\mathbf{p}}|^2}{2m}, \quad (20.2)$$

where $\hat{\mathbf{p}}$ is the momentum operator, and so on.

We have, in essence, already made use of these relations in the preceding chapter in obtaining the Schrödinger equation. If quantum-mechanical operators are connected with each other by the ordinary relations of classical mechanics, then it is sufficient to obtain the expression for one operator in order to construct subsequently the total system of operators of quantum mechanics. The limiting transition to classical mechanics as $\hbar \rightarrow 0$ will automatically be ensured, provided the initial operator is correctly chosen, taking into account this condition. Such an approach appears to be quite reasonable, although not strict. In what follows another, more consistent method for the construction of operators will be presented.

One can choose as initial operators the coordinate operator and the momentum operator.

The coordinate operator \hat{r} amounts to multiplication by this variable, as does every operator corresponding to an independent variable, i.e.

$$\hat{x} = x; \quad \hat{y} = y; \quad \hat{z} = z. \quad (20.3)$$

To establish the form of the momentum operator \hat{p} , use can be made of the fact that the free particle is described by the Schrödinger equation (6.5),

$$-\frac{\hbar^2}{2m} \nabla^2 \psi = E \psi.$$

On the other hand, by virtue of what was said above, this equation can be written as

$$\frac{1}{2m} (\hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2) \psi = E \psi.$$

Hence it follows that the operators $\hat{p}_x, \hat{p}_y, \hat{p}_z$ can be chosen in the form

$$\hat{p}_x = \frac{\hbar}{i} \frac{\partial}{\partial x}; \quad \hat{p}_y = \frac{\hbar}{i} \frac{\partial}{\partial y}; \quad \hat{p}_z = \frac{\hbar}{i} \frac{\partial}{\partial z}; \quad \hat{p} = \frac{\hbar}{i} \nabla. \quad (20.4)$$

Thus the momentum component operator amounts to differentiation with respect to the corresponding coordinate. The factor i ensures the hermiticity of the operator \hat{p} . Before considering the construction of operators which correspond to quantum-mechanical variables by a more consistent method we shall consider two questions of principle: the question of the meaning of the eigenfunctions of operators, and the question of the possibility of simultaneous measurement of two quantum-mechanical quantities.

§21. The wave function and the probability of the results of measurements

Let \hat{F} represent a certain quantum-mechanical operator for which one can write

$$\hat{F} \psi_n = F_n \psi_n.$$

For definiteness we assume that the operator \hat{F} has a discrete spectrum of eigenvalues F_n and that to each of these there corresponds an eigenfunction ψ_n (the spectrum is non-degenerate). Since the eigenfunctions ψ_n form a complete system of functions, the wave function ψ can be expanded in a

series

$$\psi = \sum_n c_n \psi_n. \quad (21.1)$$

On the basis of the principle of superposition we can conclude that the state of the system described by the wave function ψ can be written in the form of a superposition of the states with definite values F_n of the physical quantity F .

The amplitude c_m in the expansion (21.1) shows the weight with which the state ψ_m is represented in the state ψ . In other words, the amplitude c_m characterizes the probability that a value equal to F_m will be found when measurements of the quantity F are carried out on the system in the state with wave function ψ . In quantum mechanics it is assumed that this probability is equal to the square of the modulus of the amplitude in the expansion, $|c_m|^2$. Thus, if we want to find the probability of finding the value F_m for the physical quantity F when measurements are carried out on the system in the state ψ , the wave function must be expanded in terms of the eigenfunctions of the operator \hat{F} . The square of the modulus of the corresponding amplitude in the expansion, $|c_m|^2$, gives the probability sought. If the quantity F changes continuously (continuous spectrum), then one can speak of the probability that in a measurement one will obtain a value of F lying in the interval between F and $F+dF$. The corresponding probability is given by the expression

$$dW = |c(F)|^2 dF. \quad (21.2)$$

Thus, in expanding ψ in terms of plane waves (see §3) the square of the modulus of the corresponding amplitude in the expansion gives the probability that in a measurement a certain given value of the momentum will be obtained.

The probabilities of the measurements of given values of the quantity F , which are defined in the way shown above, satisfy the relations

$$\sum_n |c_n|^2 = 1, \quad \int |c(F)|^2 dF = 1 \quad (21.3)$$

(under the condition that the wave function ψ is quadratically integrable, while the eigenfunctions of the operator \hat{F} are normalized by the condition (18.7) of (18.12)).

Let us, as an example, prove the last of these relations. Making use of

(19.5) and (19.6), we obtain

$$\begin{aligned} \int c^*(F) c(F) dF &= \int c^*(F) dF \int \psi_F^*(x) \psi(x) dV = \\ &= \int \psi(x) dV \int c^*(F) \psi_F^*(x) dF = \int \psi(x) \psi^*(x) dV = 1. \end{aligned} \quad (21.4)$$

The whole set of amplitudes c_n (or $c(F)$ in the case of a continuous spectrum) determines the wave function ψ completely. Hence the definition of the amplitudes in the expansion of the wave function in terms of the eigenfunctions of an arbitrary operator is equivalent to the definition of the wave function itself.

In this connection the following terminology is often used. The wave function $\psi(x)$ is said to be a wave function given in the coordinate representation (x -representation); the whole set of all amplitudes $c(F)$ is called the wave function in the F -representation. In this sense the relations (19.5) and (19.6) must be considered as completely symmetric. The relation (19.5) expresses the expansion of the wave function ψ , taken in the coordinate representation, in terms of the eigenfunctions $\psi_F(x)$ of the operator \hat{F} which is also taken in the x -representation. The amplitudes of the expansion $c(F)$ represent the wave function in the F -representation. On the other hand, the relation (19.6) expresses the expansion of the wave function $c(F)$, taken in the F -representation, in terms of the functions $\psi_F^*(x)$ which have the meaning of the eigenfunctions of the coordinate operator taken in the F -representation (see (48.19)). The amplitudes of the expansion $\psi(x)$ represent the wave function in the x -representation. We shall say also that a certain operator \hat{D} is given in the F -representation if it acts on a function given in the F -representation, for example, $\hat{D}c(F) = b(F)$. From this point of view the statement formulated above that $|c(F)|^2 dF$ is equal to the probability of observing the system in the state with a given value of F becomes almost obvious. Indeed, $|\psi(x)|^2 dx$ is the probability that the coordinate of the particle lies in the interval dx . In view of the equivalence of the x -representation and F -representation, $|c(F)|^2 dF$ is naturally interpreted as the probability that a measurement of F will lead to a value which lies in the interval between F and $F+dF$.

§22. Mean values

Let us assume that the state of a system is described by a wave function $\psi(x)$ which is not an eigenfunction of the operator \hat{F} corresponding to the quantum-mechanical quantity F . As we have already explained above, this

means that in the given state the quantity F has no definite value. In measurements carried out on the system one can obtain, with a certain probability, any eigenvalue F_n . In this connection it is natural to try to find the mean value of the quantity F in the given state. We understand the mean, as always, to be the mathematical expectation (the arithmetic mean) of the given quantity.

Let us consider an ensemble, i.e. a large number of completely identical samples of a system. Each of these systems is described by one and the same wave function ψ . We carry out the measurement of the quantity F in each of the systems. The mean value obtained from the entirety of these measurements will be called the mean value of the quantity F . According to the general formulae of the theory of probability (see ch. 1 of Part III) we can write

$$\bar{F} = \sum_n W_n F_n, \quad (22.1)$$

where W_n is the probability of obtaining the eigenvalue F_n in measuring the quantity F . Making use of the expressions for the probabilities W_n which we have obtained in the preceding section, we have for the case of the discrete spectrum

$$\bar{F} = \sum_n |c_n|^2 F_n \quad (22.2)$$

or, if the operator \hat{F} possesses a continuous spectrum,

$$\bar{F} = \int |c(F)|^2 F dF. \quad (22.3)$$

These formulae can be transformed in such a way that, instead of the amplitudes of the expansion of the wave function in terms of the eigenfunctions of the operator \hat{F} , they contain directly the wave function $\psi(x)$ (i.e. they can be transformed into the coordinate representation or any other representation). For concreteness we assume that the operator \hat{F} possesses a discrete spectrum (in the case of a continuous spectrum the transformation formulae are derived in an analogous way). Making use of the expression (19.2) for the amplitudes, we obtain

$$\bar{F} = \sum_n c_n^* c_n F_n = \sum_n c_n F_n \int \psi_n(x) \psi^*(x) dV.$$

Since the eigenfunctions $\psi_n(x)$ satisfy the equation

$$\hat{F} \psi_n(x) = F_n \psi_n(x),$$

the last relation can be rewritten in the form

$$\bar{F} = \sum_n c_n \int \psi^* \hat{F} \psi_n dV = \int dV \psi^* \hat{F} \sum_n c_n \psi_n.$$

Taking into account (21.1), we obtain finally

$$\bar{F} = \int \psi^* \hat{F} \psi dV. \quad (22.4)$$

We note that this expression must be written in a somewhat more general form, if the wave function ψ is not normalized to unity. In this case

$$\bar{F} = \frac{\int \psi^* \hat{F} \psi dV}{\int \psi^* \psi dV}. \quad (22.5)$$

If the wave function ψ is an eigenfunction of the operator \hat{F}

$$\hat{F}\psi = F_m \psi,$$

then the quantity F has a definite value equal to the eigenvalue F_m . In this case, as was to be expected, the mean value of the quantity F is the same as this eigenvalue, $\bar{F} = F_m$.

The relation (22.4) can be the starting point in choosing the operator which corresponds to a given physical quantity. It then follows immediately that in the coordinate representation the coordinate operator amounts to multiplication by this coordinate.

Indeed, proceeding from the physical meaning of the wave function, we can write the expression for the mean value of the coordinate in the form

$$\bar{x} = \int |\psi|^2 x dV = \int \psi^* x \psi dV. \quad (22.6)$$

Comparing this expression with (22.4), we see that the coordinate operator \hat{x} is multiplication by the coordinate x . In an analogous way, if we have an arbitrary function of coordinates $U(x, y, z)$, then its mean value is given by the expression

$$\overline{U(x, y, z)} = \int |\psi|^2 U(x, y, z) dV = \int \psi^* U \psi dV. \quad (22.7)$$

It follows from this that the operator of an arbitrary function of coordinates, taken in the coordinate representation, is multiplication by the function. This, of course, corresponds to the statement we made earlier. In general the operator corresponding to a physical quantity F in its own F -representation is multiplication by the quantity F . This general statement is easily explained in the same way as we have done in the example of the coordinate. The mean value of the quantity F , obtained by means of the function $c(F)$, i.e. by

means of the wave function in the F -representation, is given by the formula

$$\bar{F} = \int |c(F)|^2 F dF = \int c^*(F) F c(F) dF.$$

On the other hand, the general expression for the mean in terms of the operator \hat{F} , taken in the F -representation, must have the form

$$\bar{F} = \int c^*(F) \hat{F} c(F) dF.$$

Comparing these expressions, we see that the operator \hat{F} in its own representation is just multiplication by F .

§23. Commutation of operators

One of the most important problems arising in quantum mechanics is that of the possibility of the simultaneous measurement of values of physical quantities of a given quantum-mechanical system.

In order that two quantities F and R may have sharp values in a state described by a wave function $\psi_n(x)$, it is obvious that this wave function must be an eigenfunction of both the operators \hat{F} and \hat{R} , i.e. that the following two equations must simultaneously be satisfied:

$$\begin{aligned}\hat{F}\psi_n(x) &= F\psi_n(x), \\ \hat{R}\psi_n(x) &= R\psi_n(x).\end{aligned}\tag{23.1}$$

We operate on the first equation with the operator \hat{R} , and on the second with the operator \hat{F} :

$$\begin{aligned}\hat{R}\hat{F}\psi_n &= \hat{R}F\psi_n = FR\psi_n, \\ \hat{F}\hat{R}\psi_n &= \hat{F}R\psi_n = RF\psi_n,\end{aligned}$$

The right-hand sides of these equations are equal and, consequently, the left-hand sides are also equal, i.e.

$$\hat{R}\hat{F}\psi_n = \hat{F}\hat{R}\psi_n$$

or

$$(\hat{R}\hat{F} - \hat{F}\hat{R})\psi_n = 0.\tag{23.2}$$

If the general eigenfunctions ψ_n form a complete system of functions, then an arbitrary wave function ψ can be expanded in this system of functions. Operating on the function with the commutator $\hat{R}\hat{F} - \hat{F}\hat{R}$, we obtain,

obviously,

$$(\hat{R}\hat{F} - \hat{F}\hat{R})\psi = \sum c_n(\hat{R}\hat{F} - \hat{F}\hat{R})\psi_n = 0. \quad (23.3)$$

The last equation can be written symbolically in the form

$$\hat{R}\hat{F} - \hat{F}\hat{R} = 0. \quad (23.4)$$

Thus we have proved that if two quantum-mechanical quantities can have sharp values simultaneously, then the corresponding operators must commute. Of course, if these quantities simultaneously have sharp values only in certain particular states (so that the common eigenfunctions ψ_n do not form a complete system of functions), then the corresponding operators do not commute (see, for example, §30).

The converse of the theorem can also be proved: if two operators \hat{F} and \hat{R} commute, they will have common eigenfunctions. To prove this we operate on the equation for the eigenfunctions of the operator \hat{F} with the operator \hat{R} . We make use of the fact that the operators \hat{F} and \hat{R} commute with each other. Then we obtain

$$\hat{F}(\hat{R}\psi) = \hat{F}\hat{R}\psi.$$

We see that the function $\psi' = \hat{R}\psi$ is also an eigenfunction of the operator \hat{F} corresponding to the eigenvalue F . If there is no degeneracy, then the function ψ' describes the same state as the function ψ , and, consequently, can differ from ψ only by the constant factor R , i.e.

$$\hat{R}\psi = R\psi.$$

Thus we have proved that the function ψ will simultaneously be an eigenfunction of the operators \hat{F} and \hat{R} . The proof is easily generalized to the case where there is degeneracy. However, in this case not every eigenfunction of the operator \hat{F} or \hat{R} will simultaneously be an eigenfunction of both operators. Nevertheless, for commuting operators one can always construct a complete system of common eigenfunctions.

Summing up what has been said above, we can say that if two quantum-mechanical quantities there correspond commuting operators, then these quantities will simultaneously have sharp values; but if the operators do not commute, then these quantities, generally speaking, cannot simultaneously have sharp values, with the exception of certain particular cases (see §30).

Let us illustrate this by a concrete example. The coordinate operator \hat{x} and the operator corresponding to the momentum component, \hat{p}_x , can be chosen as an example of non-commuting operators. The corresponding quantities x and p_x , as we know, do not simultaneously (i.e. in one and the same state)

have sharp values. On the contrary, the coordinate operator and a momentum component operator corresponding to an orthogonal coordinate, for example \hat{x} and \hat{p}_y , commute with each other. The corresponding quantities x and p_y are simultaneously measurable. The coordinate in a given direction and momentum components in a direction orthogonal to the given direction can simultaneously have sharp values.

We can now define more precisely the concept 'state of a system' in quantum mechanics. A state of a system is given if the wave function describing this system is given. However, in no circumstances can we measure directly the wave function itself. Only the square of its modulus, interpreted as a probability, has a physical meaning. A way out of this apparent contradiction is as follows. When we say that a state of a system is given, then this means that the value of a definite set of quantum-mechanical quantities is given. This set of quantities, the definition of which completely determines the state of the system, is called a complete set of quantum-mechanical quantities. In classical physics, in order to define the state of a system at an arbitrary instant of time, we have to give the values of all the generalized momenta and generalized coordinates at that instant of time. If the classical system has n degrees of freedom, we have to give the values of $2n$ variables. For a microsystem, i.e. a system described by quantum mechanics, it is obvious that the complete set cannot include both the momenta and coordinates of the particles, because these quantities have not simultaneously sharp values. To define the state of a system in quantum mechanics it is sufficient to give only the coordinates of the particle or only its momenta or, in general, any set of independent quantities which are simultaneously measurable and whose number is equal to the number of degrees of freedom of the system. Then the wave function describing the given state of the system will be an eigenfunction of the operators of the quantities which enter into the complete set corresponding to the given eigenvalues.

For example, if the system possesses three degrees of freedom, then the momentum components p_x, p_y, p_z can be chosen as the quantities which form the complete set. The corresponding wave function has the form (2.12).

States characterized by a complete set of quantities defined at a given instant of time are said to be completely specified or 'pure' states. These states are unambiguously described by a corresponding wave function. At a given instant of time this wave function is chosen as the eigenfunction of the operators of all the quantities which enter into the complete set. We note that we shall also obtain a 'pure' state in the case where the wave function corresponding to this state is represented by a certain superposition of eigenfunctions, for example by a superposition of plane waves (3.3). We obtain

information on the development of the system in time by solving the Schrödinger equation (6.8) with given initial conditions, thus defining the wave function at subsequent instants of time.

It should be noted that as well as 'pure' states one sometimes has to deal with so-called 'mixed' states (see § 89). In these states the wave function of the system is not an eigenfunction. One can speak only of the probability P_n of the realization of one or other 'pure' state φ_n .

If we are interested in the probability of measuring the F_m th value of the quantity F , then in the 'pure' state $\psi(x)$ this probability is determined by the square of the modulus $|c_m|^2$ of the corresponding amplitude in the expansion of the function ψ in terms of the eigenfunctions $\psi_k(x)$ of the operator \hat{F}

$$\begin{aligned}\psi(x) &= \sum_k c_k \psi_k(x), \\ W(F_m) &= |c_m|^2 = \left| \int \psi(x) \psi_m^*(x) dV \right|^2.\end{aligned}\quad (23.5)$$

If the system is in a 'mixed' state, then in order to obtain the required probability we have to expand the wave function $\varphi_n(x)$ in terms of the functions $\psi_k(x)$

$$\varphi_n(x) = \sum_k c_{kn} \psi_k(x),$$

where

$$c_{kn} = \int \varphi_n(x) \psi_k^*(x) dV. \quad (23.6)$$

The probability of observing the F_m th value of the quantity F in the state φ_n is given by the square of the modulus of the expansion amplitude c_{mn} . In its turn the state $\varphi_n(x)$ is realized with a probability P_n . Thus, finally, according to the theorem of the multiplication of probabilities we obtain

$$W'(F_m) = \sum_n P_n |c_{mn}|^2. \quad (23.7)$$

In order to compare the results obtained, we write the wave function ψ in the form of a superposition of the functions $\varphi_n(x)$

$$\psi(x) = \sum_n b_n \varphi_n(x).$$

Then, as is easily seen from (23.5) and (23.6),

$$c_m = \sum_n b_n c_{mn}. \quad (23.8)$$

Substituting this value into the expression (23.5), we get

$$W(F_m) = \left| \sum_n b_n c_{mn} \right|^2 = \sum_n |b_n|^2 |c_{mn}|^2 + \frac{1}{2} \sum_{n \neq k} b_n b_k c_{mn} c_{mk}, \quad (23.9)$$

where

$$\sum_n |b_n|^2 = \sum_n P_n = 1.$$

We see that the expression obtained differs from the result given by formula (23.7) by the presence of the double sum which expresses interference between the states. In the case of 'mixed' states there is no such interference.

The above reasoning can be generalized directly to the case of a continuous spectrum.

§24. Heisenberg inequalities

We have in the preceding section found the conditions under which the simultaneous measurement of two physical quantities is possible. We now assume that two physical quantities F and R do not simultaneously have sharp values. Then the operators \hat{F} and \hat{R} corresponding to these quantities do not commute with each other. We assume that the following relation holds:

$$\hat{F}\hat{R} - \hat{R}\hat{F} = i\hat{B}, \quad (24.1)$$

where \hat{B} is a certain Hermitian operator.

It is of interest to determine in a general form the minimum possible value of the product of fluctuations of given quantities. We choose the mean square deviations (dispersions) $\overline{\Delta F^2}$ and $\overline{\Delta R^2}$, where

$$\Delta F = F - \bar{F},$$

$$\Delta R = R - \bar{R},$$

as the measure characterizing the deviations of the individual results of measuring the quantities F and R from their mean values. Correspondingly, for the mean square deviations we have

$$\begin{aligned} \overline{\Delta F^2} &= \overline{(F - \bar{F})^2} = \overline{F^2} - \bar{F}^2, \\ \overline{\Delta R^2} &= \overline{(R - \bar{R})^2} = \overline{R^2} - \bar{R}^2. \end{aligned} \quad (24.2)$$

Without affecting the generality of the argument we can set $\bar{F} = 0$ and

$\bar{R} = 0$ (in other words, we can understand F and R to be the deviation of these quantities from their mean value).

Let us consider the integral

$$J(\alpha) = \int |(\alpha\hat{F} - i\hat{R})\psi|^2 dV. \quad (24.3)$$

Here ψ is the wave function, the integration is carried out over the entire region of variation of the independent variables, and α is an arbitrary real parameter. The integral (24.3) is not negative: $J(\alpha) \geq 0$. We rewrite it in the form

$$J(\alpha) = \int (\alpha\hat{F} - i\hat{R})\psi \cdot (\alpha\hat{F}^* + i\hat{R}^*)\psi^* dV.$$

Making use of the self-conjugate property of the operators \hat{F} and \hat{R} , we obtain

$$\begin{aligned} J(\alpha) &= \int \psi^* (\alpha\hat{F} + i\hat{R}) (\alpha\hat{F} - i\hat{R}) \psi dV = \\ &= \int \psi^* (\alpha^2 \hat{F}^2 - \alpha i(\hat{F}\hat{R} - \hat{R}\hat{F}) + \hat{R}^2) \psi dV. \end{aligned}$$

Taking into account (24.1) and using expression (22.4) for the mean value, we have

$$J(\alpha) = \alpha^2 \overline{F^2} + \alpha \bar{B} + \overline{R^2} = \alpha^2 \overline{\Delta F^2} + \alpha \bar{B} + \overline{\Delta R^2}.$$

The condition for this trinomial quadratic in α to be negative can be written in the form

$$4 \overline{\Delta F^2} \overline{\Delta R^2} \geq \bar{B}^2. \quad (24.4)$$

or

$$(\overline{\Delta F^2} \overline{\Delta R^2})^{\frac{1}{2}} \geq \frac{1}{2} |\bar{B}|. \quad (24.5)$$

Formula (24.5) gives the relation which we have sought between the uncertainties ΔF and ΔR . It establishes the minimum possible value of the product of these errors.

Let us consider a particular case, taking as the quantities F and R respectively p_x and x . Then it follows from (20.3) and (20.4) that $\bar{B} = -\hbar$, and we have

$$(\overline{\Delta p_x^2})^{\frac{1}{2}} (\overline{\Delta x^2})^{\frac{1}{2}} \geq \frac{1}{2} \hbar. \quad (24.6)$$

Thus the uncertainty relation (24.5) is of a general character. The uncertainty relation for the coordinate and momentum is a particular case of the relation (24.5).

Conjugate quantum-mechanical quantities cannot be measured simulta-

neously. The minimum uncertainties in their values in a simultaneous measurement are connected with the quantity \bar{B} . On the contrary, mutually commuting quantum-mechanical quantities, for which $\hat{B} = 0$, can be measured simultaneously with an arbitrary degree of accuracy.

§25. Poisson brackets

In §20 we have considered one of the possible methods of finding the operators describing physical quantities. This problem was considered in a more consistent way by Dirac. He assumed that in quantum mechanics, as in classical mechanics, the concept of Poisson brackets can be introduced*. Thus, if to two classical quantities F, R there corresponds the Poisson bracket

$$[F, R] = - \sum_i \left(\frac{\partial F}{\partial p_i} \frac{\partial R}{\partial q_i} - \frac{\partial F}{\partial q_i} \frac{\partial R}{\partial p_i} \right),$$

then to the operators \hat{F}, \hat{R} describing these quantities there corresponds the quantum Poisson bracket $[\hat{F}, \hat{R}]$. Further it was assumed that the properties of the quantum Poisson brackets correspond exactly to those of the classical Poisson brackets except that for the quantum brackets the consecutive order of the two factors is important. We write down the properties of the Poisson brackets:

$$[\hat{F}, \hat{R}] = - [\hat{R}, \hat{F}], \quad (25.1)$$

$$[\hat{F}, C] = 0, \quad (25.2)$$

where C is a constant.

$$[\hat{F}_1 + \hat{F}_2, \hat{R}] = [\hat{F}_1, \hat{R}] + [\hat{F}_2, \hat{R}], \quad (25.3)$$

$$[\hat{F}, \hat{R}_1 + \hat{R}_2] = [\hat{F}, \hat{R}_1] + [\hat{F}, \hat{R}_2], \quad (25.4)$$

$$[\hat{F}_1 \hat{F}_2, \hat{R}] = [\hat{F}_1, \hat{R}] \hat{F}_2 + \hat{F}_1 [\hat{F}_2, \hat{R}], \quad (25.5)$$

$$[\hat{F}, \hat{R}_1 \hat{R}_2] = [\hat{F}, \hat{R}_1] \hat{R}_2 + \hat{R}_1 [\hat{F}, \hat{R}_2], \quad (25.6)$$

$$[\hat{F}_1, [\hat{F}_2, \hat{F}_3]] + [\hat{F}_3, [\hat{F}_1, \hat{F}_2]] + [\hat{F}_2, [\hat{F}_3, \hat{F}_1]] = 0. \quad (25.7)$$

The choice of Poisson brackets as the basis for the construction of a

* For a discussion of Poisson brackets in classical mechanics see L.D.Landau and E.M.Lifshitz, *Mechanics* (Pergamon Press, Oxford, 1960); H.Goldstein, *Classical mechanics* (Addison-Wesley, Cambridge, Mass., 1950).

system of quantum-mechanical operators is associated with the fact that, as we shall see, they can be expressed directly in terms of the commutators of the corresponding operators. The last combination of operators is the basis for their physical interpretation.

Let us consider the Poisson bracket $[\hat{F}_1 \hat{F}_2, \hat{R}_1 \hat{R}_2]$, for the calculation of which we can use expressions (25.5) and (25.6). Correspondingly we obtain

$$\begin{aligned} [\hat{F}_1 \hat{F}_2, \hat{R}_1 \hat{R}_2] &= [\hat{F}_1, \hat{R}_1 \hat{R}_2] \hat{F}_2 + \hat{F}_1 [\hat{F}_2, \hat{R}_1 \hat{R}_2] = \\ &= [\hat{F}_1, \hat{R}_1] \hat{R}_2 \hat{F}_2 + \hat{R}_1 [\hat{F}_1, \hat{R}_2] \hat{F}_2 + \hat{F}_1 [\hat{F}_2, \hat{R}_1] \hat{R}_2 + \hat{F}_1 \hat{R}_1 [\hat{F}_2, \hat{R}_2] \end{aligned}$$

and

$$\begin{aligned} [\hat{F}_1 \hat{F}_2, \hat{R}_1 \hat{R}_2] &= [\hat{F}_1 \hat{F}_2, \hat{R}_1] \hat{R}_2 + \hat{R}_1 [\hat{F}_1 \hat{F}_2, \hat{R}_2] = \\ &= \hat{F}_1 [\hat{F}_2, \hat{R}_1] \hat{R}_2 + [\hat{F}_1, \hat{R}_1] \hat{F}_2 \hat{R}_2 + \hat{R}_1 \hat{F}_1 [\hat{F}_2, \hat{R}_2] + \hat{R}_1 [\hat{F}_1, \hat{R}_2] \hat{F}_2. \end{aligned}$$

Equating these two results, we find

$$(\hat{F}_1 \hat{R}_1 - \hat{R}_1 \hat{F}_1) [\hat{F}_2, \hat{R}_2] = [\hat{F}_1, \hat{R}_1] (\hat{F}_2 \hat{R}_2 - \hat{R}_2 \hat{F}_2).$$

Since the last equality must be satisfied identically, we have

$$\begin{aligned} [\hat{F}_1, \hat{R}_1] &= iC(\hat{F}_1 \hat{R}_1 - \hat{R}_1 \hat{F}_1), \\ [\hat{F}_2, \hat{R}_2] &= iC(\hat{F}_2 \hat{R}_2 - \hat{R}_2 \hat{F}_2), \end{aligned}$$

where C is a real constant.

The reality of C follows from the fact that the Poisson bracket of two real variables must also be real. Thus, if $\hat{F}^\dagger = \hat{F}$, $\hat{R}^\dagger = \hat{R}$, then also $[\hat{F}, \hat{R}]^\dagger = [\hat{F}, \hat{R}]$. However,

$$[\hat{F}, \hat{R}]^\dagger = -iC^*(\hat{F}\hat{R} - \hat{R}\hat{F})^\dagger = -iC^*(\hat{R}\hat{F} - \hat{F}\hat{R}) = C^*C^{-1}[\hat{F}, \hat{R}]$$

and hence $C^* = C$. It follows from the classical theory of Poisson brackets that the constant C has the dimensionality $\text{erg}^{-1} \cdot \text{sec}^{-1}$. Its numerical value can be determined only by comparing the inferences of the theory with experimental data. It turns out to be equal to \hbar^{-1} . Finally we have

$$(\hat{F}\hat{R} - \hat{R}\hat{F}) = \frac{\hbar}{i} [\hat{F}, \hat{R}]. \quad (25.8)$$

In the transition to classical mechanics (see ch. 5), i.e. for $\hbar \rightarrow 0$, the commutator $\{\hat{F}\hat{R}\}$ reduces to zero, as was to be expected. It is natural to assume if only in the simplest cases, that the Poisson brackets themselves have the same values as classical brackets. For canonically conjugated variable

coordinates and momenta in classical mechanics we have:

$$\begin{aligned} [p_i, p_k] &= 0, \\ [x_i, x_k] &= 0, \\ [p_i, x_k] &= \delta_{ik} \quad (i, k = 1, 2, 3). \end{aligned} \quad (25.9)$$

Here and in what follows we make use of the notation

$$\begin{aligned} x_1 &= x, & p_1 &= p_x, \\ x_2 &= y, & p_2 &= p_y, \\ x_3 &= z, & p_3 &= p_z. \end{aligned}$$

The same expressions can be written for the quantum operators of the coordinate and of the momentum components. Hence the commutators of the corresponding values assume the form

$$\begin{aligned} \hat{x}_i \hat{x}_k - \hat{x}_k \hat{x}_i &= 0, \\ \hat{p}_i \hat{p}_k - \hat{p}_k \hat{p}_i &= 0, \\ \hat{p}_i \hat{x}_k - \hat{x}_k \hat{p}_i &= \frac{\hbar}{i} \delta_{ik} \quad (i, k = 1, 2, 3). \end{aligned} \quad (25.10)$$

We shall make use of these equalities for the determination of the coordinate operator and the momentum operator.

§26. Coordinate and momentum operators and eigenfunctions

We begin with the establishment of the form of operators in the coordinate representation*. In this representation the wave function characterizing the state of the particle depends on its coordinates $\psi(x, y, z)$. The coordinates x, y, z are independent variables. Hence the corresponding operators, in correspondence with the conclusions of §20 and §22, reduce to multiplication by these coordinates

$$\hat{x}_i = x_i \quad (i = 1, 2, 3). \quad (26.1)$$

* See V.A.Fok, *Nachala kvantovoi mekhaniki (Principles of quantum mechanics)* (KUBUCH, 1932) p.32.

We rewrite the commutation relations (25.10) in the form

$$(\hat{p}_k x_k - x_k \hat{p}_k) \psi = \frac{\hbar}{i} \psi \quad (k = 1, 2, 3), \quad (26.2)$$

$$(\hat{p}_i x_k - x_k \hat{p}_i) \psi = 0 \quad (i \neq k), \quad (26.3)$$

$$(\hat{p}_i \hat{p}_k - \hat{p}_k \hat{p}_i) \psi = 0 \quad (i, k = 1, 2, 3). \quad (26.4)$$

Equations (26.2)–(26.4) can be satisfied by an arbitrary function ψ , setting

$$\hat{p}_k = \frac{\hbar}{i} \frac{\partial}{\partial x_k} + \frac{\partial \alpha(x_1, x_2, x_3)}{\partial x_k}, \quad (26.5)$$

where $\alpha(x_1, x_2, x_3)$ is an arbitrary real function. The reality of α is required for the hermiticity of the operator \hat{p}_k . The function α can, without loss of generality of the result, be set equal to zero. Indeed, the action of the operator (26.5) on an arbitrary function ψ transforms it into the function

$$\psi' = \frac{\hbar}{i} \frac{\partial \psi}{\partial x_k} + \frac{\partial \alpha}{\partial x_k} \psi.$$

On the other hand, if we act on the function $e^{(i/\hbar)\alpha} \psi$ with the operator $(\hbar/i) (\partial/\partial x_k)$, we then obtain the function $e^{(i/\hbar)\alpha} \psi'$. Consequently, the transition from the operator (26.5) to the operator $(\hbar/i) (\partial/\partial x_k)$ is equivalent to the transition from the wave function ψ to the function $e^{(i/\hbar)\alpha} \psi$

$$\begin{aligned} \psi &\rightarrow e^{(i/\hbar)\alpha} \psi, \\ \frac{\hbar}{i} \frac{\partial}{\partial x_k} + \frac{\partial \alpha}{\partial x_k} &\rightarrow \frac{\hbar}{i} \frac{\partial}{\partial x_k} = e^{(i/\hbar)\alpha} \left(\frac{\hbar}{i} \frac{\partial}{\partial x_k} + \frac{\partial \alpha}{\partial x_k} \right) e^{-(i/\hbar)\alpha}, \end{aligned} \quad (26.6)$$

since

$$e^{(i/\hbar)\alpha} \left(\frac{\hbar}{i} \frac{\partial}{\partial x_k} + \frac{\partial \alpha}{\partial x_k} \right) e^{-(i/\hbar)\alpha} e^{(i/\hbar)\alpha} \psi = e^{(i/\hbar)\alpha} \psi' = \frac{\hbar}{i} \frac{\partial}{\partial x_k} (e^{(i/\hbar)\alpha} \psi).$$

Wave functions, and operators, connected with each other by the transformations (26.6) have identical physical properties. This will be proved in its most general form in §46.

The operators (26.5) and $\hat{p}_k = (\hbar/i) (\partial/\partial x_k)$ ($k = x, y, z$) have the same spectrum of eigenvalues. Hence without loss of generality we can use, instead of the operators (26.5), the operators for the momentum components which have, in the coordinate representation, the form (see §20)

$$\hat{p}_x = \frac{\hbar}{i} \frac{\partial}{\partial x}; \quad \hat{p}_y = \frac{\hbar}{i} \frac{\partial}{\partial y}; \quad \hat{p}_z = \frac{\hbar}{i} \frac{\partial}{\partial z} \quad (26.7)$$

or in vector form

$$\hat{\mathbf{p}} = \frac{\hbar}{i} \nabla, \quad (26.8)$$

where ∇ is the gradient operator.

Let us now use the momentum representation, in which the wave function depends on three momentum components: p_x, p_y, p_z . The corresponding operators reduce to multiplication by the quantities p_x, p_y, p_z . The coordinate operators in this representation are found on the basis of the same commutation relations, and turn out to be equal to

$$\hat{x} = i\hbar \frac{\partial}{\partial p_x}; \quad \hat{y} = i\hbar \frac{\partial}{\partial p_y}; \quad \hat{z} = i\hbar \frac{\partial}{\partial p_z}, \quad (26.9)$$

or

$$\hat{\mathbf{r}} = i\hbar \left(\mathbf{i} \frac{\partial}{\partial p_x} + \mathbf{j} \frac{\partial}{\partial p_y} + \mathbf{k} \frac{\partial}{\partial p_z} \right) \equiv i\hbar \frac{\partial}{\partial \mathbf{p}}.$$

Making use of (26.7) it is easy to establish the commutation relations of the operator $\hat{\mathbf{p}}$ and an arbitrary function $U(x, y, z)$

$$\hat{\mathbf{p}}U - U\hat{\mathbf{p}} = \frac{\hbar}{i} \nabla U. \quad (26.10)$$

The commutation of the operator $\hat{\mathbf{r}}$ with an arbitrary function $f(p_x, p_y, p_z)$ is calculated in an analogous way

$$\hat{\mathbf{r}}f - f\hat{\mathbf{r}} = i\hbar \frac{\partial f}{\partial \mathbf{p}}. \quad (26.11)$$

The equations for the eigenfunctions and eigenvalues of the operators $\hat{p}_x, \hat{p}_y, \hat{p}_z$ are of the form

$$\frac{\hbar}{i} \frac{\partial \psi_{p_x}}{\partial x} = p_x \psi_{p_x}, \quad \frac{\hbar}{i} \frac{\partial \psi_{p_y}}{\partial y} = p_y \psi_{p_y}, \quad \frac{\hbar}{i} \frac{\partial \psi_{p_z}}{\partial z} = p_z \psi_{p_z}. \quad (26.12)$$

We write down the solution of the first equation

$$\psi_{p_x} = a(y, z) e^{(i/\hbar)p_x x},$$

where $a(y, z)$ is an arbitrary function. Analogous solutions are also obtained for the functions ψ_{p_y} and ψ_{p_z} . The functions $\psi_{p_x}, \psi_{p_y}, \psi_{p_z}$ satisfy the necessary requirements, in particular the condition of finiteness (see § 16) for real values of p_x, p_y, p_z . Thus the momentum operator has a continuous spectrum

of eigenvalues. The wave function

$$\psi_{\mathbf{p}} = A e^{(i/\hbar)(\mathbf{p} \cdot \mathbf{r})}, \quad (26.13)$$

where A is a constant, is an eigenfunction of the operators $\hat{p}_x, \hat{p}_y, \hat{p}_z$ and describes the state with given momentum \mathbf{p} . A freely moving particle can be in such a state. This conclusion is in complete agreement with the result of §2.

The constant A is defined by the normalization condition. Since the momentum operator has a continuous spectrum, its eigenfunctions are conveniently normalized to the δ -function. Let us first find the normalization coefficient A in the case of one-dimensional motion.

Setting $\int \psi_{p_x}^* \psi_{p'_x} dx = \delta(p_x - p'_x)$ and taking into account (III.5), we obtain $A = (2\pi\hbar)^{-1/2}$, so that finally

$$\psi_{p_x} = (2\pi\hbar)^{-1/2} e^{(i/\hbar)p_x x}. \quad (26.14)$$

In the three-dimensional case, for the wave function (26.13) we have correspondingly

$$\psi_{\mathbf{p}} = (2\pi\hbar)^{-3/2} e^{(i/\hbar)(\mathbf{p} \cdot \mathbf{r})}. \quad (26.15)$$

Another method of normalizing plane waves, called normalization in a 'box', sometimes turns out to be more convenient. We define the wave function in an arbitrarily large but finite volume V . As the normalization volume we choose a cube with edge L and centre at the origin. We require that at the walls of the cube the wave functions (26.12) satisfy the condition of periodicity, i.e. that at corresponding points of opposite faces the wave functions take on the same values. Under these conditions the momentum vector no longer changes continuously, but runs over a discrete set of values

$$p_x = \frac{2\pi\hbar}{L} n_x; \quad p_y = \frac{2\pi\hbar}{L} n_y; \quad p_z = \frac{2\pi\hbar}{L} n_z, \quad (26.16)$$

where n_x, n_y, n_z are positive or negative integers, including zero. Choosing the edge of the cube, L , to be sufficiently large, the spacing between neighbouring eigenvalues of the momentum vector can be made as small as one pleases. The normalization coefficient defined by the condition

$$|A|^2 \int_{L^3} |e^{(i/\hbar)\mathbf{p} \cdot \mathbf{r}}|^2 dV = 1$$

is equal to $A = L^{-3/2}$. Correspondingly the wave function for such a normali-

zation has the form

$$\psi_{\mathbf{p}} = L^{-\frac{3}{2}} e^{(i/\hbar)(\mathbf{p} \cdot \mathbf{r})} = V^{-\frac{3}{2}} e^{(i/\hbar)(\mathbf{p} \cdot \mathbf{r})}. \quad (26.17)$$

In §12 and 13 we normalized wave functions of the form (26.13), defining the probability current density \mathbf{j}_0 . As a matter of fact, in this state, according to (7.6),

$$\mathbf{j}_0 = |A|^2 \frac{\mathbf{p}}{m}. \quad (26.18)$$

Setting, for example, $A = 1$, we obtain

$$\mathbf{j}_0 = \frac{\mathbf{p}}{m} = \mathbf{v}, \quad (26.19)$$

i.e. for such a normalization the probability current density is numerically equal to the velocity of the particle. But if $A = v^{-\frac{1}{2}}$, then this corresponds to a normalization to unit probability current density.

It is easily seen that the operators $\hat{p}_x, \hat{p}_y, \hat{p}_z$ are related in a simple way to the operators of an infinitesimal translation along the x -, y - and z -axes respectively. Indeed, let us shift our system or, what is equivalent, the origin, a distance Δx along the x -axis. Then the old and new coordinates are connected by the relation

$$x' = x - \Delta x,$$

$$y' = y,$$

$$z' = z.$$

We express the function $\psi(x, y, z)$ in terms of the new coordinates x', y', z' . Confining ourselves to the first term of a series expansion we obtain

$$\begin{aligned} \psi(x, y, z) &= \psi(x' + \Delta x, y', z') = \psi(x', y', z') + \frac{\partial \psi}{\partial x'} \Delta x = \\ &= \left(1 + \Delta x \frac{\partial}{\partial x'} \right) \psi(x', y', z'). \end{aligned}$$

It is natural to call the operator $1 + \Delta x (\partial/\partial x')$ the operator of displacement by a distance Δx along the x -axis. We denote this operator by $\hat{R}_{x'}$, so that

$$\psi(x, y, z) = \hat{R}_{x'} \psi(x', y', z'). \quad (26.20)$$

We see that the displacement operator \hat{R}_x is connected with the operator of

the corresponding momentum component \hat{p}_x

$$\hat{R}_x = 1 + \frac{i}{\hbar} \Delta x \hat{p}_x. \quad (26.21)$$

The form of the momentum operator \hat{p}_x could also be obtained proceeding from the expression for the operator \hat{R}_x . *

We write down the equation for the eigenfunctions and eigenvalues of the coordinate operator in the coordinate representation

$$\hat{x} \psi_{x_0}(x) = x_0 \psi_{x_0}(x). \quad (26.22)$$

Here x_0 is a particular value of the coordinate x . The operator \hat{x} in its own representation reduces to multiplication by x . From eq. (26.22) it follows that

$$\psi_{x_0}(x) = 0 \quad \text{for } x \neq x_0.$$

Furthermore, the functions $\psi_{x_0}(x)$ must satisfy the orthogonality and normalization conditions

$$\int \psi_{x_0}^*(x) \psi_{x'_0}(x) dx = \delta(x_0 - x'_0).$$

From these relations it follows that the function $\psi_{x_0}(x)$ is of the form (see Appendix III)

$$\psi_{x_0}(x) = \delta(x - x_0). \quad (26.23)$$

The eigenfunctions of the operators \hat{y} and \hat{z} are written in an analogous way. Since the coordinate operators \hat{x} , \hat{y} , \hat{z} commute with each other, their values are simultaneously measurable. Correspondingly, if the system has three degrees of freedom, the three coordinate projections x , y , z can be chosen as quantities forming a complete set. The wave function describing the state with three defined coordinates x_0, y_0, z_0 is of the form

$$\psi_{\mathbf{r}_0}(\mathbf{r}) = \delta(x - x_0) \delta(y - y_0) \delta(z - z_0) = \delta(\mathbf{r} - \mathbf{r}_0). \quad (26.24)$$

The eigenfunction of the momentum operator in the momentum representation is written in an analogous way.

* See L.D. Landau and E.M. Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965). An analogous result also applies to the operators of the y - and z -momentum components.

§27. The Hamiltonian operator

The most important operator of quantum mechanics is the total energy operator \hat{H} . As in classical mechanics, it is made up of the kinetic energy operator and the potential energy operator. We construct, first of all, the operator of the kinetic energy of a particle. In the non-relativistic approximation, in which we are now interested, the kinetic energy is connected with the momentum of the particle by the usual relation

$$T = \frac{\mathbf{p}^2}{2m} = \frac{p_x^2 + p_y^2 + p_z^2}{2m}. \quad (27.1)$$

Replacing the momentum \mathbf{p} of the particle in this relation by the operator $\hat{\mathbf{p}}$, we obtain the operator \hat{T} which we shall call the kinetic energy operator (see also §20):

$$\hat{T} = \frac{1}{2m} \hat{\mathbf{p}}^2 = \frac{1}{2m} (\hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2) \equiv -\frac{\hbar^2}{2m} \nabla^2. \quad (27.2)$$

It is obvious that the kinetic energy operator commutes with the momentum operator.

We now consider the total energy operator \hat{H} . Since the potential energy depends only on the coordinates x, y, z , the corresponding operator in the coordinate representation is simply the function $U(x, y, z)$. Consequently we have

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla^2 + U(x, y, z). \quad (27.3)$$

Since the total energy operator in formula (27.3) is expressed in terms of the momentum operator (but not the velocity operator), it represents the quantum-mechanical Hamiltonian operator which is usually just called the Hamiltonian. The expression for the Hamiltonian can easily be generalized to the case where the particle moves in non-stationary external fields. Then

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla^2 + U(\mathbf{r}, t), \quad (27.4)$$

where $U(\mathbf{r}, t)$ is the so-called force function, connected with the force which acts on the particle by the relation

$$\mathbf{f} = -\nabla U.$$

The formulae found for the Hamiltonian operator are inapplicable in the case

of the motion of a particle in a field of force which depends on its velocity. An example of such a case is the motion of a charged particle in a magnetic field.

To obtain the Hamiltonian operator in this case we make use of general rules. We write down the Hamiltonian function of classical mechanics for particles moving in an electromagnetic field. According to (41.4) of Part I, we have

$$H = \frac{1}{2m} \left(\mathbf{p} - \frac{e}{c} \mathbf{A} \right)^2 + e\varphi, \quad (27.5)$$

where the vector \mathbf{p} is the generalized momentum of the particle, \mathbf{A} and φ are the vector and scalar potentials, and e is the charge of the particle. According to the general rule, we replace the Hamiltonian function in formula (27.5) by the Hamiltonian operator, and the generalized momentum by the momentum operator. The vector potential and scalar potential, which depend only on the coordinates and time, can be left unchanged, since in the coordinate representation the application of the corresponding operators amounts to multiplication by these functions. We then find

$$\hat{H} = \frac{1}{2m} \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right)^2 + e\varphi. \quad (27.6)$$

By means of the Hamiltonian operator found above the basic equation of quantum mechanics, the Schrödinger equation, can be written in the form

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H}\psi. \quad (27.7)$$

The operator form of notation of the Schrödinger equation has a very general character and is suitable for the description of the motion of a particle in arbitrary stationary or non-stationary fields. In particular, in such a form it is valid for the case of the motion of a particle in an electromagnetic field. Just as for the classical Hamiltonian function, the Hamiltonian can be transformed to an arbitrary curvilinear system of coordinates. For this one need only transform the differential Laplacian operator ∇^2 to this system. Depending on the symmetry of the force field it is convenient to choose a system of curvilinear coordinates in which the expression for the potential energy of the particle assumes the simplest form. In particular, as we shall see in §35, it is often convenient to write the Hamiltonian operator in spherical coordinates.

The Hamiltonian operator of a system of particles can be constructed according to the same scheme which has already been successfully applied to the case of one particle. Namely, one has to write the classical expression for

the Hamiltonian function and then replace all quantities involved in it by their quantum-mechanical operators.

The classical expression for the Hamiltonian of a system of N particles has the form

$$H = \sum_{k=1}^N \frac{\mathbf{p}_k^2}{2m_k} + \sum_{k=1}^N U_k(\mathbf{r}_k) + U_{\text{int}}, \quad (27.8)$$

where \mathbf{p}_k , m_k and $U_k(\mathbf{r}_k)$ are respectively the momentum, mass and potential energy of the k th particle in the external field; U_{int} is the potential energy of interaction of the particles.

We obtain the Hamiltonian operator if we replace the momenta of the particles by the corresponding operators $\hat{\mathbf{p}}_k$, where the index k denotes differentiation with respect to the coordinates of the k th particle. After this replacement we obtain the Schrödinger equation. It has the form

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H} \psi, \quad \hat{H} = \sum_{k=1}^N -\frac{\hbar^2}{2m_k} \nabla_k^2 + \sum_{k=1}^N U_k(\mathbf{r}_k) + U_{\text{int}}. \quad (27.9)$$

The expression for the Hamiltonian operator for a system of charged particles in an external electromagnetic field is generalized directly from (27.6).

§28. Stationary states

We assume that the Hamiltonian of the system does not depend explicitly on the time. In this case it is possible to separate the variables in the Schrödinger equation (27.7). We have already made use of this in §6. However, we can now analyze more profoundly the solution obtained.

We seek the solution of the Schrödinger equation (27.7) in the form

$$\psi(x, t) = \chi(t) \psi(x), \quad (28.1)$$

where we understand x to be the entire set of coordinates on which the wave function depends.

Substituting this expression into (27.7), we obtain

$$i\hbar \frac{d\chi(t)}{dt} \psi(x) = \chi(t) \hat{H} \psi(x).$$

Separating variables in this equation gives:

$$i\hbar \frac{d\chi}{dt} \frac{1}{\chi} = \frac{\hat{H}\psi(x)}{\psi(x)}.$$

The expression on the left-hand side of the equation can depend only on the time t , while the expression on the right-hand side can depend only on the coordinates of the system. It follows from the equality of these expressions that each of them is equal to one and the same constant which we shall denote by E . We then obtain

$$\chi(t) = Ce^{-(i/\hbar)Et}, \quad \hat{H}\psi(x) = E\psi(x),$$

where C is an arbitrary constant.

We see that the constant E has the meaning of an eigenvalue of the operator \hat{H} , i.e. determines the possible values of the energy of the system, and the function $\psi(x)$ describes a state with given energy.

The Hamiltonian operator can possess a discrete as well as a continuous spectrum, as we have seen in the examples already discussed. One also often encounters a mixed spectrum, i.e. a discrete spectrum in one energy interval, and a continuous spectrum in another.

Assuming for definiteness that the operator \hat{H} possesses a discrete spectrum, we write the wave functions (28.1)

$$\psi_n(x, t) = \psi_n(x) e^{-(i/\hbar)E_n t}. \quad (28.2)$$

The states of a system described by wave functions of the type (28.2) are said to be stationary. The wave functions of stationary states depend harmonically on the time with frequencies $\omega_n = E_n/\hbar$. As we have already noted in §6, in a stationary state the density of the probability of finding a particle at a given point of space does not depend on the time. Indeed,

$$W_n(x, t) = |\psi_n(x, t)|^2.$$

Substituting the expression (28.2) for the wave function, we find

$$W_n(x, t) = |\psi_n(x)|^2 = W_n(x, 0). \quad (28.3)$$

This statement can easily be generalized. The probability of observing the eigenvalue F_k in the stationary state $\psi_n(x, t)$ does not depend on the time. According to our general rules (see §21), in order to obtain the probability to be determined we must expand the wave function $\psi_n(x, t)$ in terms of the eigenfunctions ψ_{F_k} of the operator \hat{F} and take the square of the modulus of

the corresponding expansion amplitude c_k . According to formula (19.2)

$$c_k(t) = \int \psi_n(x, t) \psi_{F_k}^*(x) dV = e^{-(i\hbar)E_n t} \int \psi_n(x) \psi_{F_k}^*(x) dV.$$

The corresponding probability $W(F_k, t)$ is equal to

$$W(F_k, t) = |c_k(t)|^2 = \left| \int \psi_n(x) \psi_{F_k}^*(x) dV \right|^2 = W(F_k, 0). \quad (28.4)$$

An arbitrary solution of the Schrödinger equation $\psi(x, t)$ can be expanded in terms of the wave functions (28.2). This function $\psi(x, t)$ describes a state in which the energy of the system has no sharp value.

§29. Integral form of the Schrödinger equation

The Schrödinger differential equation can also be represented by an integral equation. In a number of cases this latter formulation has a number of advantages both from the theoretical standpoint and from the point of view of purely mathematical convenience. The theoretical advantage of the integral representation of the equations of quantum mechanics is closely associated with the development of Feynman's ideas in quantum field theory (see ch. 14).

In §58 we shall dwell in detail on the advantages of approximate methods of solving the Schrödinger equation in integral form.

Let us consider a particle with Hamiltonian \hat{H} depending, in general, on the time. At the initial instant of time let the wave function

$$\psi_0 = \psi(\mathbf{r}_1, t_1) \quad (29.1)$$

be defined. The wave function satisfies the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H}\psi. \quad (29.2)$$

The wave function of the particle satisfying eq. (29.2) for the boundary condition (29.1) at the instant of time $t_2 > t_1$ can be written in the form

$$\psi(\mathbf{r}_2, t_2) = \int K(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1) \psi(\mathbf{r}_1, t_1) d\mathbf{r}_1. \quad (29.3)$$

The function $K(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1)$ is the Green's function of eq. (29.2) (see §19). The interpretation of formula (29.3) is obvious: the Green's function represents the amplitude for the transition of the particle from the initial state with wave function $\psi(\mathbf{r}_1, t_1)$ into the state with wave function $\psi(\mathbf{r}_2, t_2)$, where $t_2 > t_1$.

Since formula (29.3) defines the wave function only for $t_2 > t_1$, the Green's function can be predetermined by the requirement

$$K(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1) = 0 \quad \text{for } t_2 < t_1.$$

For the relation (29.3) to be equivalent to (29.2) and (29.1), the Green's function must satisfy the equation

$$\left[i\hbar \frac{\partial}{\partial t_2} - \hat{H}(\mathbf{r}_2, t_2) \right] K(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1) = i\hbar \delta(\mathbf{r}_2 - \mathbf{r}_1) \delta(t_2 - t_1). \quad (29.4)$$

Indeed, for $t_2 > t_1$ the Green's function satisfies the equation

$$\left[i\hbar \frac{\partial}{\partial t_2} - \hat{H}(\mathbf{r}_2, t_2) \right] K(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1) = 0. \quad (29.5)$$

Operating on both sides of (29.3) with the operator $i\hbar(\partial/\partial t_2) - \hat{H}$ for $t_2 > t_1$, and taking into account eq. (29.5), we arrive at an identity. It is easily seen that the Green's function defined by eq. (29.4) satisfies the initial condition (29.1) for $t_2 = t_1$, if (29.4) is integrated over the infinitesimal interval $2\Delta t \rightarrow 0$ about the instant t_1 . We then have

$$\begin{aligned} \int_{t_1 - \Delta t}^{t_1 + \Delta t} \left[i\hbar \frac{\partial}{\partial t_2} - \hat{H}(\mathbf{r}_2, t_2) \right] K(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1) dt_2 = \\ = i\hbar \delta(\mathbf{r}_2 - \mathbf{r}_1) \int_{t_1 - \Delta t}^{t_1 + \Delta t} \delta(t_2 - t_1) dt_2 = i\hbar \delta(\mathbf{r}_2 - \mathbf{r}_1). \end{aligned}$$

Evidently we have

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \int_{t_1 - \Delta t}^{t_1 + \Delta t} \hat{H}(\mathbf{r}_2, t_2) K(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1) dt_2 = 0, \\ \lim_{\Delta t \rightarrow 0} \int_{t_1 - \Delta t}^{t_1 + \Delta t} i\hbar \frac{\partial}{\partial t_2} K(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1) dt_2 = i\hbar K(\mathbf{r}_2, t_1; \mathbf{r}_1, t_1). \end{aligned}$$

Hence

$$K(\mathbf{r}_2, t_1; \mathbf{r}_1, t_1) = \delta(\mathbf{r}_2 - \mathbf{r}_1). \quad (29.6)$$

Thus if the Green's function satisfies eq. (29.4), then (29.3) represents the solution of the Schrödinger equation with the corresponding initial conditions (the Cauchy problem).

In other words, if the transition amplitude is known, then the wave function is also known. On the other hand, the transition amplitude has certain important features which make it in many respects more convenient, and as completely characteristic of the system as the wave function.

Let us first consider the case where the Hamiltonian \hat{H}_0 of the system does not depend explicitly on the time. Then one can find the general relation between the wave functions of the stationary states $\psi_n(\mathbf{r}, t)$, the energy levels of the system, and the transition amplitude. We shall denote the latter by K_0 for a system with Hamiltonian \hat{H}_0 . The amplitude must satisfy the equation

$$\left[i\hbar \frac{\partial}{\partial t_2} - \hat{H}_0(\mathbf{r}_2) \right] K_0(2, 1) = i\hbar \delta(\mathbf{r}_2 - \mathbf{r}_1) \delta(t_2 - t_1). \quad (29.7)$$

If \hat{H}_0 does not depend explicitly on the time, then the wave function satisfying eq. (29.2) can be written in the form

$$\psi_n = U_n(\mathbf{r}) \exp [-(i/\hbar) E_n t],$$

where the ψ_n form a complete set of orthonormal functions. By virtue of this property of the ψ_n , one can always write the expansion

$$K_0(2, 1) \equiv K_0(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1) = \sum C_n(\mathbf{r}_1, t_1) \psi_n(\mathbf{r}_2, t_2) \theta(t_2 - t_1), \quad (29.8)$$

where $\theta(t_2 - t_1)$ is the step function

$$\theta(x) = \begin{cases} 1 & (x > 0) \\ 0 & (x < 0) \end{cases}. \quad (29.9)$$

The behaviour of K_0 for $t_2 < t_1$ is taken into account by means of the θ -function. Substituting the expansions (29.8) into (29.4), we find

$$\begin{aligned} \left[i\hbar \frac{\partial}{\partial t_2} - \hat{H}_0(\mathbf{r}_2) \right] \sum C_n(\mathbf{r}_1, t_1) \psi_n(\mathbf{r}_2, t_2) \theta(t_2 - t_1) = \\ = i\hbar \sum C_n \psi_n \frac{d\theta(t_2 - t_1)}{dt_2} + \sum C_n \theta(t_2 - t_1) \left[i\hbar \frac{\partial}{\partial t_2} - \hat{H}_0(\mathbf{r}_2) \right] \psi_n = \\ = i\hbar \sum C_n \psi_n \delta(t_2 - t_1). \end{aligned}$$

On the other hand using (29.7),

$$\left[i\hbar \frac{\partial}{\partial t_2} - \hat{H}_0(\mathbf{r}_2) \right] \sum C_n \psi_n \theta(t_2 - t_1) = i\hbar \delta(\mathbf{r}_2 - \mathbf{r}_1) \delta(t_2 - t_1).$$

Hence $\sum C_n \psi_n = \delta(\mathbf{r}_2 - \mathbf{r}_1)$. But for ψ_n the normalization condition holds:

$$\sum \psi_n^*(\mathbf{r}_1) \psi_n(\mathbf{r}_1) = \delta(\mathbf{r}_2 - \mathbf{r}_1).$$

Hence for the coefficient C_n we find $C_n = \psi_n^*$, and, finally, we obtain

$$K_0(2, 1) = \theta(t_2 - t_1) \sum_n U_n(\mathbf{r}_2) U_n(\mathbf{r}_1) \exp [-(i/\hbar) E_n(t_2 - t_1)] . \quad (29.10)$$

The summation goes over into integration for a continuous spectrum. As an example, let us find the explicit expression for the Green's function of a free particle. In this case

$$\psi = (2\pi\hbar)^{-\frac{3}{2}} \exp(i\mathbf{p} \cdot \mathbf{r}) \exp \left(-\frac{ip^2 t}{2m\hbar} \right) ,$$

so that

$$\begin{aligned} K_0(2, 1) &= \theta(t_2 - t_1) (2\pi\hbar)^{-3} \int \exp[-i\mathbf{p} \cdot (\mathbf{r}_1 - \mathbf{r}_2)] \exp \left(-\frac{ip^2(t_2 - t_1)}{2m\hbar} \right) d\mathbf{p} = \\ &= \frac{\theta(t_2 - t_1)}{[2\pi i\hbar(t_2 - t_1)/m]^{\frac{3}{2}}} \exp \left(\frac{im\hbar |\mathbf{r}_2 - \mathbf{r}_1|^2}{2(t_2 - t_1)} \right) . \end{aligned}$$

The transition amplitude $K_0(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1)$, or more briefly $K_0(2, 1)$, has the following important properties:

- (1) The transition amplitude depends only on the difference $t_2 - t_1$, as is seen from formula (29.10): $K_0(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1) = K_0(\mathbf{r}_2, \mathbf{r}_1; t_2 - t_1)$.
- (2) Owing to this the transition amplitude $K_0(\mathbf{r}_3, t_3; \mathbf{r}_1, t_1) = K_0(\mathbf{r}_3, \mathbf{r}_1; t_3 - t_1)$ can be written in the form

$$K_0(\mathbf{r}_3, \mathbf{r}_1; t_3 - t_1) = \int K_0(\mathbf{r}_3, \mathbf{r}_2; t_3 - t_2) K_0(\mathbf{r}_2, \mathbf{r}_1; t_2 - t_1) d\mathbf{r}_2 . \quad (29.11)$$

This means that the transition can be considered as a set of successive transitions $(1 \rightarrow 2), (2 \rightarrow 3)$ for all possible states 2.

Formula (29.11) expresses the principle of superposition. Its proof is elementary:

$$\begin{aligned} \psi(\mathbf{r}_3, t_3) &= \int K_0(\mathbf{r}_3, \mathbf{r}_2; t_3 - t_2) \psi(\mathbf{r}_2, t_2) d\mathbf{r}_2 = \\ &= \iint K_0(\mathbf{r}_3, \mathbf{r}_2; t_3 - t_2) K_0(\mathbf{r}_2, \mathbf{r}_1; t_2 - t_1) \psi(\mathbf{r}_1, t_1) d\mathbf{r}_2 d\mathbf{r}_1 , \\ \psi(\mathbf{r}_3, t_3) &= \int K_0(\mathbf{r}_3, \mathbf{r}_1; t_3 - t_1) \psi(\mathbf{r}_1, t_1) d\mathbf{r}_1 . \end{aligned}$$

Comparison of these formulae gives (29.11) immediately.

- (3) The Fourier component of the transition amplitude K_0 defines the spectrum of eigenvalues of the energy of the system. Let us find the Fourier

component of the function $K_0(\mathbf{r}_2, \mathbf{r}_1; t_2 - t_1)$

$$\begin{aligned} K_0(\mathbf{r}_2, \mathbf{r}_1; \omega) &= \int K_0(\mathbf{r}_2, \mathbf{r}_1; t_2 - t_1) \exp[i\omega(t_2 - t_1)] d(t_2 - t_1) = \\ &= \sum U_n(\mathbf{r}_2) U_n^*(\mathbf{r}_1) \int \exp[-(i/\hbar)E_n(t_2 - t_1)] \exp[i\omega(t_2 - t_1)] \times \\ &\times \theta(t_2 - t_1) d(t_2 - t_1) = \\ &= \sum U_n(\mathbf{r}_2) U_n^*(\mathbf{r}_1) I, \end{aligned}$$

where

$$I = \int \theta(t_2 - t_1) \exp[-(i/\hbar)E_n(t_2 - t_1)] \exp[i\omega(t_2 - t_1)] d(t_2 - t_1).$$

The θ -function can be written in the form of a contour integral

$$\theta(x) = \lim_{\gamma \rightarrow 0} \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{i\alpha x}}{\alpha - i\gamma} d\alpha. \quad (29.12)$$

The contour of integration for $x > 0$ is shown in fig. V.8.

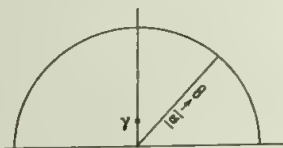


Fig. V.8

Then for I we find

$$\begin{aligned} I &= \lim_{\gamma \rightarrow 0} \frac{1}{2\pi i} \int \frac{d\alpha}{\alpha - i\gamma} \int \exp[-i(\omega_n - \omega - \alpha)\tau] d\tau = \\ &= \lim_{\gamma \rightarrow 0} \frac{1}{i} \int \frac{d\alpha}{\alpha - i\gamma} \delta(\omega_n - \omega - \alpha) = \lim_{\gamma \rightarrow 0} \frac{i}{\omega - \omega_n - i\gamma}. \end{aligned} \quad (29.13)$$

Hence for $K_0(\mathbf{r}_2, \mathbf{r}_1; \omega)$ we obtain

$$K_0(\mathbf{r}_2, \mathbf{r}_1; \omega) = \lim_{\gamma \rightarrow 0} \sum U_n(\mathbf{r}_2) U_n^*(\mathbf{r}_1) \frac{1}{\omega - (E_n/\hbar) - i\gamma}. \quad (29.14)$$

We see that the poles of the Fourier component of the transition amplitude, $\omega = E_n/\hbar$, correspond to the energy eigenvalues. Thus knowing the transition amplitude K_0 one can find directly the spectrum of energies E_n .

Let us turn to the general case of a Hamiltonian depending on time. It can usually be written in the form of a sum $\hat{H} = \hat{H}_0(\mathbf{r}) + U(\mathbf{r}, t)$, where \hat{H}_0 does

not depend on time. $U(\mathbf{r}, t)$ often represents a variable external field acting on the particle. In this case the Green's function K satisfies the equation

$$\left[i\hbar \frac{\partial}{\partial t_2} - \hat{H}_0 - U \right] K(2, 1) = i\hbar \delta(\mathbf{r}_2 - \mathbf{r}_1) \delta(t_2 - t_1) \quad (29.15)$$

and reduces to zero for $t_2 < t_1$,

$$K(2, 1) = 0 \quad \text{for } t_2 < t_1. \quad (29.16)$$

The differential equation (29.15) for the Green's function can be compared to the integral equation

$$K(2, 1) = K_0(2, 1) - \frac{i}{\hbar} \int K_0(2, 3) U(3) K(3, 1) d^4x_3 \quad (29.17)$$

where $d^4x = dx dy dz dt$. In this integral equation $K_0(2, 1)$ is considered to be a known function, and $K_0(2, 3) U(3)$ is its kernel. This can easily be seen by acting on eq. (29.17) with the operator $[i\hbar(\partial/\partial t_2) - \hat{H}_0(\mathbf{r}_2)]$. Then, taking into account (29.7), we obtain

$$\left[i\hbar \frac{\partial}{\partial t_2} - \hat{H}_0(\mathbf{r}_2) \right] K(2, 1) = i\hbar \delta(t_2 - t_1) \delta(\mathbf{r}_2 - \mathbf{r}_1) + U(2) K(2, 1).$$

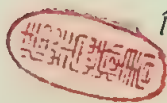
Thus we again arrive at eq. (29.15).

The initial condition (29.16) is contained in (29.17) since $K_0(2, 1) = 0$ for $t_2 < t_1$. The integral form of the equation for the transition amplitude (29.7) is especially convenient because it allows one to obtain $K(2, 1)$ in the form of a series of successive approximations (see §58).

§30. The eigenvalues and eigenfunctions of the angular momentum operator and of the operator of the square of the angular momentum

Let us now form some operators which play important roles in our subsequent considerations – the operators of the angular momentum components and of the square of the angular momentum. Replacing mechanical quantities in the classical definition of angular momentum by quantum-mechanical operators according to general rule, we find

$$\begin{aligned} \hat{l}_x &= y\hat{p}_z - z\hat{p}_y = \frac{\hbar}{i} \left(y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right), \\ \hat{l}_y &= z\hat{p}_x - x\hat{p}_z = \frac{\hbar}{i} \left(z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right), \\ \hat{l}_z &= x\hat{p}_y - y\hat{p}_x = \frac{\hbar}{i} \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right). \end{aligned} \quad (30.1)$$



We shall call the set of operators \hat{l}_x , \hat{l}_y , and \hat{l}_z the angular momentum operator $\hat{\mathbf{l}}$. This quantity possesses all the properties of angular momentum. In particular, as we shall show below, it obeys the same conservation laws as angular momentum in classical mechanics.

Further, we construct the operator of the square of the angular momentum

$$\hat{l}^2 = \hat{l}_x^2 + \hat{l}_y^2 + \hat{l}_z^2. \quad (30.2)$$

Let us consider the commutation relations for these operators. We first of all note that not all of the operators of the different angular momentum components commute with each other. Let us calculate, for example, the commutator $\hat{l}_x \hat{l}_y - \hat{l}_y \hat{l}_x$. Making use of expressions (30.1) we have

$$\begin{aligned} \hat{l}_x \hat{l}_y &= -\hbar^2 \left(y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) \left(z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) = \\ &= -\hbar^2 \left(y \frac{\partial}{\partial x} + yz \frac{\partial^2}{\partial z \partial x} - xy \frac{\partial^2}{\partial z^2} - z^2 \frac{\partial^2}{\partial y \partial x} + xz \frac{\partial^2}{\partial y \partial z} \right). \end{aligned}$$

On the other hand, interchanging the operators we find

$$\begin{aligned} \hat{l}_y \hat{l}_x &= -\hbar^2 \left(z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) \left(y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) = \\ &= -\hbar^2 \left(zy \frac{\partial^2}{\partial x \partial z} - z^2 \frac{\partial^2}{\partial x \partial y} - xy \frac{\partial^2}{\partial z^2} + x \frac{\partial}{\partial y} + xz \frac{\partial^2}{\partial z \partial y} \right). \end{aligned}$$

Subtracting the lower equation from the upper, we obtain finally

$$\hat{l}_x \hat{l}_y - \hat{l}_y \hat{l}_x = \hbar^2 \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) = i\hbar \hat{l}_z. \quad (30.3)$$

Carrying out the cyclic permutation of the coordinates x, y, z we obtain two more relations:

$$\hat{l}_y \hat{l}_z - \hat{l}_z \hat{l}_y = i\hbar \hat{l}_x, \quad \hat{l}_z \hat{l}_x - \hat{l}_x \hat{l}_z = i\hbar \hat{l}_y. \quad (30.3')$$

From the relations (30.3) it follows that the components of the angular momentum of a particle l_x, l_y, l_z cannot simultaneously have sharp values. An exception to this is the state when the angular momentum is equal to zero, so that then $\hat{l}_x = \hat{l}_y = \hat{l}_z = 0$. On the other hand the angular momentum component operators \hat{l}_x , \hat{l}_y , and \hat{l}_z do commute with the operator of the

square of the angular momentum $\hat{\mathbf{I}}^2$, i.e. the following relations hold:

$$\begin{aligned}\hat{I}_x \hat{\mathbf{I}}^2 - \hat{\mathbf{I}}^2 \hat{I}_x &= 0, \\ \hat{I}_y \hat{\mathbf{I}}^2 - \hat{\mathbf{I}}^2 \hat{I}_y &= 0, \\ \hat{I}_z \hat{\mathbf{I}}^2 - \hat{\mathbf{I}}^2 \hat{I}_z &= 0.\end{aligned}\tag{30.4}$$

These relations are easily proved by means of (30.3). Let us prove, for example, the first of these. From relation (30.3), multiplying on the right and on the left by \hat{I}_y , we have

$$\begin{aligned}\hat{I}_x \hat{I}_y^2 &= \hat{I}_y \hat{I}_x \hat{I}_y + i\hbar \hat{I}_z \hat{I}_y, \\ \hat{I}_y^2 \hat{I}_x &= \hat{I}_y \hat{I}_x \hat{I}_y - i\hbar \hat{I}_y \hat{I}_z.\end{aligned}$$

Subtracting the second relation from the first, we obtain

$$\hat{I}_x \hat{I}_y^2 - \hat{I}_y^2 \hat{I}_x = i\hbar (\hat{I}_z \hat{I}_y + \hat{I}_y \hat{I}_z).$$

Analogously,

$$\hat{I}_x \hat{I}_z^2 - \hat{I}_z^2 \hat{I}_x = -i\hbar (\hat{I}_z \hat{I}_y + \hat{I}_y \hat{I}_z).$$

Also taking into account the fact that $\hat{I}_x \hat{I}_x^2 - \hat{I}_x^2 \hat{I}_x = 0$ and adding up the equalities obtained, we find

$$\hat{I}_x \hat{\mathbf{I}}^2 - \hat{\mathbf{I}}^2 \hat{I}_x = 0.$$

The two remaining relations (30.4) are proved in the same way. From these relations it follows that the square of the total angular momentum and one of its projections onto an arbitrary axis can simultaneously have definite values.

We note that commutation relations analogous to (30.3), (30.3') also hold for the angular momentum operator and the coordinate operator, and the angular momentum operator and momentum operator. Omitting the simple proof, we write down the two relations

$$\begin{aligned}\hat{I}_x \hat{y} - \hat{y} \hat{I}_x &= i\hbar \hat{z}, \\ \hat{I}_x \hat{p}_y - \hat{p}_y \hat{I}_x &= i\hbar \hat{p}_z.\end{aligned}\tag{30.5}$$

The remaining four relations are obtained by the cyclic permutation of the indices. Relations (30.3), (30.4) and (30.5) are the same as the corresponding classical expressions if, of course, we pass from commutators to the classical Poisson brackets.

Further, let us determine the possible values of the angular momentum projection onto an arbitrarily chosen direction in space and the possible values of the square of the angular momentum (i.e. the eigenvalues of the

corresponding operators). In solving the equations for the eigenfunctions and eigenvalues it is convenient to use spherical coordinates.

We carry out the transition from the Cartesian coordinates x, y, z to the variables r, ϑ, φ in formulae (30.1) and (30.2) according to the ordinary rules of replacement of variables. Omitting these elementary calculations, we simply given the result

$$\hat{l}_z = \frac{\hbar}{i} \frac{\partial}{\partial \varphi}, \quad (30.6)$$

$$\hat{l}_x = \frac{\hbar}{i} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} + \cotan \vartheta \cos \varphi \frac{\partial}{\partial \varphi} \right), \quad (30.7)$$

$$\hat{l}_y = \frac{\hbar}{i} \left(\cos \varphi \frac{\partial}{\partial \vartheta} - \cotan \vartheta \sin \varphi \frac{\partial}{\partial \varphi} \right), \quad (30.8)$$

$$\hat{l}^2 = -\hbar^2 \left[\frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2} \right] = -\hbar^2 \nabla_{\vartheta, \varphi}^2, \quad (30.9)$$

where $\nabla_{\vartheta, \varphi}^2$ is the angular part of the Laplacian in spherical coordinates.

Choosing an arbitrary direction in space as the z -axis, we define the eigenfunctions and eigenvalues of the operator of the component of angular momentum in this direction. The equation for the eigenfunctions and eigenvalues of the operator \hat{l}_z is of the form

$$\frac{\hbar}{i} \frac{\partial \psi}{\partial \varphi} = l_z \psi. \quad (30.10)$$

The solution of this equation is

$$\psi = \psi(r, \vartheta) \exp(i l_z \varphi / \hbar) \quad (30.11)$$

where $\psi(r, \vartheta)$ is an arbitrary function.

The wave function which is the solution of eq. (30.10) must satisfy the condition of single-valuedness. Since φ is a cyclic variable varying from 0 to 2π , the condition of single-valuedness is written in the form $\psi(\varphi) = \psi(\varphi + 2\pi)$ or

$$\exp[(i/\hbar) l_z \varphi] = \exp[(i/\hbar) l_z (\varphi + 2\pi)].$$

This last condition is fulfilled if $l_z = m\hbar$, where m is a positive or negative integer (including zero). In what follows it will be called the magnetic quantum number.

Since the z -axis is not specified by any physical condition, the same result also holds for the operators \hat{l}_x and \hat{l}_y .

Thus the angular momentum component along an arbitrarily chosen direction in space takes on integer (in units of \hbar) values. For a sharp value of the projection l_z the two other projections have no well defined value. This means that if in a state with given l_z the values of the projections l_x and l_y are measured, then any possible value may be found for them.

The eigenfunction of the operator \hat{l}_z depending on the angle φ and normalized to unity by the condition

$$\int_0^{2\pi} \psi_m^*(\varphi) \psi_{m'}(\varphi) d\varphi = \delta_{mm'},$$

has the form

$$\psi_m(\varphi) = (2\pi)^{-1/2} e^{im\varphi}. \quad (30.12)$$

Let us now determine the eigenvalues and eigenfunctions of the operator of the square of the angular momentum, \hat{l}^2

$$\hat{l}^2 \psi = l^2 \psi. \quad (30.13)$$

Substituting into (30.13) the expression for \hat{l}^2 given by formula (30.9), we obtain the equation

$$\frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial \psi}{\partial \vartheta} \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2 \psi}{\partial \varphi^2} + \frac{l^2}{\hbar^2} \psi = 0. \quad (30.14)$$

The equation for the eigenfunctions of the operator \hat{l}^2 is the well-known equation for the spherical harmonic functions*.

Equation (30.14) only has solutions satisfying the standard conditions formulated in § 16 for the values $l^2/\hbar^2 = l(l+1)$, where l is a positive integer (including zero). The quantum number l is called the orbital angular momentum quantum number. Thus the operator of the square of the angular momentum has a discrete spectrum of eigenvalues

$$l^2 = \hbar^2 l(l+1). \quad (30.15)$$

The solution of eq. (30.14) for the eigenfunctions of the operator of the

* See, for example, V.I. Smirnov, *A course of higher mathematics*, Vol. III (Pergamon Press, Oxford, 1964) and V.A. Fok, *Nachala kvantovoi mekhaniki (Principles of quantum mechanics)* (KUBUCH, 1932) p. 118.

square of the angular momentum is of the form

$$\begin{aligned}\psi_{lm}(\vartheta, \varphi) &= Y_{lm}(\vartheta, \varphi) = \\ &= (-1)^k \left(\frac{(l-|m|)! (2l+1)}{(l+|m|)! 4\pi} \right)^{\frac{1}{2}} P_l^m(\cos \vartheta) e^{im\varphi}, \quad (30.16)\end{aligned}$$

where m is an integer taking on the values $m = 0, \pm 1, \pm 2, \dots, \pm l$; $k = m$ for $m \geq 0$ and $k = 0$ for $m < 0$.

We denote by P_l^m the associated Legendre polynomial

$$P_l^m = (1 - \xi^2)^{\frac{1}{2}|m|} \frac{d^{|m|}}{d\xi^{|m|}} P_l(\xi) = \frac{1}{2^l l!} (1 - \xi^2)^{\frac{1}{2}|m|} \frac{d^{|m|+l}}{d\xi^{|m|+l}} (\xi^2 - 1)^l. \quad (30.17)$$

The constant factor in formula (30.16) is defined by the condition of normalization of the function Y_{lm} to unity.

$$\int_0^\pi \int_0^{2\pi} Y_{lm}^*(\vartheta, \varphi) Y_{l'm'}(\vartheta, \varphi) \sin \vartheta d\vartheta d\varphi = \delta_{ll'} \delta_{mm'}. \quad (30.18)$$

From formulae (30.15) and (30.16) it follows that to each eigenvalue of the square of the angular momentum there correspond $2l+1$ eigenfunctions Y_{lm} (differing in the number m). Thus the eigenvalues of the square of the angular momentum are degenerate. The meaning of this degeneracy, and consequently of the number m , is easily understood. We act on the wave function Y_{lm} with the operator \hat{l}_z . We then obtain

$$\hat{l}_z Y_{lm}(\vartheta, \varphi) = \hbar m Y_{lm}(\vartheta, \varphi). \quad (30.19)$$

We see that the wave function Y_{lm} is simultaneously an eigenfunction of the operators \hat{l}_z and \hat{l}^2 . Hence it is clear that the quantum number m involved in (30.16) characterizes the value of the angular momentum projection onto the z -axis in a given state, and the wave function Y_{lm} describes a state with a given total angular momentum and a given projection on the z -axis.

Summarizing we can say that the value of the total angular momentum is defined according to formula (30.15) by the orbital angular momentum quantum number l running over a sequence of integer values. For a fixed value of the square of the angular momentum the projection of the angular momentum onto an arbitrarily oriented z -axis can take on $2l+1$ values from $-l$ up to $+l$ (in units of \hbar). Any other values of this angular momentum projection for a given l are impossible. Since the z -axis is oriented quite arbitrarily, it is natural that the angular momentum projections onto the x -axis and y -axis for a given l also take on values from $-l$ up to $+l$. For $l = 0$ the

angular momentum projection onto any axis is also equal to zero. This is the only state in which the angular momentum projections onto different axes simultaneously have sharp values. In this case the function Y_{lm} ($l = 0$) reduces to a constant which is an eigenfunction of all the operators $\hat{l}_x, \hat{l}_y, \hat{l}_z$.

We note that the eigenvalue of the square of the total angular momentum $l^2 = \hbar^2 l(l+1)$ is always larger than the square of the maximum projection of the angular momentum which is equal to $\hbar^2 l^2$. If these quantities were the same, then this would mean that in a state in which the angular momentum projection onto a certain axis has its maximum value the other two projections would be equal to zero. This is, however, impossible, since for a sharp value of one of the angular momentum projections the other two cannot have well defined values, not even zero.

Finally, we shall show that the angular momentum operator is related to the operator of an infinitesimal rotation of the system about the origin. Let us rotate the coordinate system through a small angle $\delta\varphi$ about, for example, the z -axis. The old and new coordinates of a point are connected by the relation

$$\begin{aligned}x' &= x + y \delta\varphi, & x &= x' - y' \delta\varphi, \\y' &= -x \delta\varphi + y, & y &= x' \delta\varphi + y', \\z' &= z, & z &= z' .\end{aligned}$$

Consequently, upon rotation the wave function $\psi(x, y, z)$ expressed in terms of the new variables has the form

$$\begin{aligned}\psi(x, y, z) &= \psi(x' - y' \delta\varphi, y' + x' \delta\varphi, z') = \\&= \psi(x', y', z') - y' \delta\varphi \frac{\partial \psi}{\partial x'} + x' \delta\varphi \frac{\partial \psi}{\partial y'} = \\&= \left[1 + \delta\varphi \left(x' \frac{\partial}{\partial y'} - y' \frac{\partial}{\partial x'} \right) \right] \psi(x', y', z) = \\&= \left(1 + \frac{i}{\hbar} \delta\varphi \hat{l}_z \right) \psi(x', y', z') = \hat{W}_z \psi(x', y', z') .\end{aligned}$$

It is natural to call the operator \hat{W} the rotation operator. We found the operator \hat{W}_z of the rotation through a small angle $\delta\varphi$ about the z -axis to be connected with the operator \hat{l}_z by the relation

$$\hat{W}_z = 1 + \frac{i}{\hbar} \delta\varphi \hat{l}_z . \quad (30.20)$$

Such a relation also holds, of course, for any other axis.

§31. Differentiation of operators with respect to time

We now construct the operator \hat{F} corresponding to the derivative with respect to time of the quantum-mechanical quantity described by the operator \hat{F} . It is clear that the ordinary definition of the derivative of a function is inapplicable to the quantum-mechanical quantity described by the operator \hat{F} . To define the notion of the derivative we again make use of the analogy with classical mechanics. As is known, in classical mechanics the derivative with respect to time of a mechanical quantity F can be expressed in terms of the classical Poisson bracket

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + [H, F] ,$$

where H is the Hamiltonian.

Passing from classical quantities to quantum-mechanical operators and from the classical Poisson bracket to the quantum one, we obtain the expression for the operator \hat{F}

$$\hat{F} = \frac{\partial \hat{F}}{\partial t} + [\hat{H}, \hat{F}] . \quad (31.1)$$

If the operator \hat{F} does not depend explicitly on time, then the operator \hat{F} has the form

$$\hat{F} = [\hat{H}, \hat{F}] = \frac{i}{\hbar} (\hat{H}\hat{F} - \hat{F}\hat{H}) . \quad (31.2)$$

The expressions for the derivative of the sum \hat{F} and product \hat{L} of two operators \hat{D} and \hat{R}

$$\hat{F} = \hat{D} + \hat{R} , \quad (31.3)$$

$$\hat{L} = \hat{D}\hat{R} + \hat{D}\hat{R} \quad (31.4)$$

follow immediately from the properties of the quantum Poisson brackets.

By means of formula (31.1) for the derivative of the quantum operator one can find the expression for the derivative with respect to time of the mean value of the quantity F .

Differentiating the expression (22.4) for the mean, we find

$$\dot{\bar{F}} = \int \frac{\partial \psi^*}{\partial t} \hat{F} \psi dV + \int \psi^* \frac{\partial \hat{F}}{\partial t} \psi dV + \int \psi^* \hat{F} \frac{\partial \psi}{\partial t} dV .$$

We express the derivatives $\partial \psi / \partial t$ and $\partial \psi^* / \partial t$ in terms of the wave functions

by means of the Schrödinger equation and the equation which is conjugate to it. We then have

$$\begin{aligned}\dot{\bar{F}} &= \int \psi^* \frac{\partial \hat{F}}{\partial t} \psi \, dV - \frac{i}{\hbar} \int \psi^* \hat{F} (\hat{H} \psi) \, dV + \frac{i}{\hbar} \int (\hat{H}^* \psi^*) \hat{F} \psi \, dV, \\ \int (\hat{H}^* \psi^*) \hat{F} \psi \, dV &= \int (\hat{F} \psi) \hat{H}^* \psi^* \, dV,\end{aligned}$$

because the integral does not change when the integrands are exchanged. It follows from the Hermitian property of the operator \hat{H} that

$$\int (\hat{F} \psi) \hat{H}^* \psi^* \, dV = \int \psi^* \hat{H} \hat{F} \psi \, dV.$$

We finally obtain

$$\dot{\bar{F}} = \int \psi^* \left(\frac{\partial \hat{F}}{\partial t} + \frac{i}{\hbar} (\hat{H} \hat{F} - \hat{F} \hat{H}) \right) \psi \, dV. \quad (31.5)$$

Comparing the above expression with the definition of the mean of the derivative $\dot{\bar{F}}$, we arrive at the important equality $\dot{\bar{F}} = \overline{\dot{F}}$.

As an example let us define the operators \hat{x} and \hat{p}_x . Since the coordinate operator and momentum operator do not depend explicitly on the time, we have

$$\hat{x} = [\hat{H}, \hat{x}], \quad \hat{p}_x = [\hat{H}, \hat{p}_x]. \quad (31.6)$$

In such a form the operator equations (31.6) are analogous to the classical Hamilton equations. We evaluate the commutators on the right-hand sides of the equations (31.6), assuming that the Hamiltonian has the form

$$\hat{H} = \frac{1}{2m} (\hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2) + U(x, y, z, t).$$

Taking into account that the coordinate and momentum operators are

$$\hat{x} = x, \quad \hat{p}_x = \frac{\hbar}{i} \frac{\partial}{\partial x},$$

we obtain

$$[\hat{H}, \hat{x}] = \frac{i}{2m\hbar} (\hat{p}_x^2 x - x \hat{p}_x^2),$$

since x and $U(x, y, z, t)$ commute.

Calculating the commutator of the operators \hat{p}_x^2 and x , we find

$$\begin{aligned}\hat{p}_x^2 x - x \hat{p}_x^2 &= -\hbar^2 \left(\frac{\partial^2}{\partial x^2} x - x \frac{\partial^2}{\partial x^2} \right) = \\ &= -\hbar^2 \left(x \frac{\partial^2}{\partial x^2} + 2 \frac{\partial}{\partial x} - x \frac{\partial^2}{\partial x^2} \right) = \\ &= -2\hbar^2 \frac{\partial}{\partial x} = -2i\hbar \hat{p}_x,\end{aligned}$$

We finally obtain

$$\hat{x} = [\hat{H}, \hat{x}] = \frac{1}{m} \hat{p}_x. \quad (31.7)$$

We see that the velocity operator \hat{x} is connected with the momentum operator \hat{p}_x by the same relation as in classical mechanics. We find the operator \hat{p}_x by

$$[\hat{H}, \hat{p}_x] = \frac{i}{\hbar} (U \hat{p}_x - \hat{p}_x U) = -\frac{\partial U}{\partial x}.$$

Thus we have

$$\hat{p}_x = -\frac{\partial U}{\partial x}. \quad (31.8)$$

We have obtained the operator equation of motion in the form of Newton's equation. Equations (31.7) and (31.8) can also be written for the mean values of the corresponding quantities

$$\bar{\ddot{x}} = \dot{\bar{x}} = \frac{1}{m} \bar{p}_x, \quad \bar{\dot{p}}_x = \dot{\bar{p}}_x = -\frac{\partial \bar{U}}{\partial x}. \quad (31.9)$$

These last relations are called the Ehrenfest theorems. Expressing $\dot{\bar{p}}_x$ in terms of $\bar{\ddot{x}}$, we find

$$m \bar{\ddot{x}} = -\frac{\partial \bar{U}}{\partial x}. \quad (31.10)$$

In such a form this equation is very close in appearance to Newton's equation of classical mechanics.

§ 32. Constants of the motion

Suppose the operator \hat{F} does not depend explicitly on the time and commutes with the Hamiltonian \hat{H} . In this case, according to (31.2), the operator corresponding to the derivative with respect to time is equal to zero, and from relation (31.5) it follows that the mean value of the quantity F does not change with time

$$\dot{\bar{F}} = 0. \quad (32.1)$$

The probability that, in measuring F , we shall obtain a possible value F_n is also constant in time. Indeed, this probability is given by the square of the modulus of the coefficient of expansion $|c_n(t)|^2$ of the wave function $\psi(x, t)$ describing the state of the system at the instant of time t in terms of the eigenfunctions of the operator \hat{F} . Since, however, the operator \hat{F} commutes with the operator \hat{H} , both operators have the common eigenfunctions

$$\psi_n(x, t) = \psi_n(x) \exp[(-i/\hbar) E_n t]$$

(see §23). The expansion of $\psi(x, t)$ in terms of the eigenfunctions of the operator \hat{F} can be written in the form

$$\psi(x, t) = \sum_n c_n(0) \exp[(-i/\hbar) E_n t] \psi_n(x) = \sum_n c_n(t) \psi_n(x). \quad (32.2)$$

Consequently,

$$|c_n(t)|^2 = |c_n(0)|^2 = \text{const.}$$

In quantum mechanics such quantities, as in classical mechanics, are usually called constants of the motion. From the above it is clear that a quantum-mechanical quantity is a constant of the motion if: (1) its operator does not depend explicitly on the time, (2) this operator commutes with the Hamiltonian.

Knowing the operators corresponding to different quantum-mechanical quantities and the Hamiltonian, one can find the conservation laws.

Finding conservation laws in quantum mechanics is as important for the study of the motion of a system as in classical mechanics. As in classical mechanics*, the laws of conservation of momentum and angular momentum are closely associated with the properties of homogeneity and isotropy of space. Thus from the isotropy of space it follows that the Hamiltonian of a

* See, for example, L.D. Landau and E.M. Lifshitz, *Mechanics* (Pergamon Press, Oxford, 1960).

closed system or of a system in a centrally symmetric force field must not change when an arbitrary infinitesimal rotation is made. Mathematically this is expressed by the fact that the Hamiltonian \hat{H} must commute with the rotation operator \hat{W} . But, as we know (see §30), the operator corresponding to rotation through a small angle about a certain axis (for example the z-axis) is related in a simple way to the operator of the component of angular momentum along this axis. Therefore a consequence of the commutation of the operator \hat{W}_z with the Hamiltonian \hat{H} is the commutation of the operator \hat{l}_z with the Hamiltonian, hence the law of conservation of this quantity follows. The fact that we have considered rotation only through a small angle is not important, since a rotation through a finite angle can be resolved into a succession of small rotations.

Thus we see that the conservation of angular momentum is associated with the isotropy of space.

Similarly it is easily seen that momentum conservation is associated with the homogeneity of space. Indeed, from the homogeneity of space it follows that the displacement operator must not change the Hamiltonian of a closed system, i.e. it must commute with the Hamiltonian. However, since the displacement operator \hat{R} is related to the operator of the corresponding momentum component (see §26), we arrive immediately at the momentum conservation law.

The law of conservation of energy in a closed system or a system in a stationary external field can be associated with the arbitrariness of the choice of the zero of time (homogeneity in time). This means that the laws of motion of the system must not depend on the choice of the zero of time.

We introduce the operator corresponding to the translation over a small time interval δt , $\hat{\mathcal{V}}(\delta t)$, defined by the relation

$$\hat{\mathcal{V}}(\delta t) \psi(x, t) = \psi(x, t + \delta t). \quad (32.3)$$

Expanding the function $\psi(x, t + \delta t)$ in a series in terms of the small interval δt and confining ourselves to terms of the first order of small quantities, we obtain

$$\hat{\mathcal{V}}(\delta t) \psi(x, t) = \left(1 + \delta t \frac{\partial}{\partial t} \right) \psi(x, t).$$

Hence it follows that the operator $\hat{\mathcal{V}}(\delta t)$ is of the form

$$\hat{\mathcal{V}}(\delta t) = 1 + \delta t \frac{\partial}{\partial t}. \quad (32.4)$$

The requirement of the independence of the laws of motion of the system

of the choice of zero time is expressed by the commutation of the operator $\hat{\mathcal{V}}(\delta t)$ with the Hamiltonian of the system

$$\hat{\mathcal{V}}(\delta t)\hat{H} = \hat{H}\hat{\mathcal{V}}(\delta t). \quad (32.5)$$

Using the expression (32.4) for $\hat{\mathcal{V}}(\delta t)$, we can rewrite relation (32.5) in the form

$$\frac{\partial \hat{H}}{\partial t} = 0. \quad (32.6)$$

But eq. (32.6) just expresses the energy conservation law. Indeed, the operator \hat{H} commutes with itself, and the condition $\hat{H} = 0$, denoting the energy conservation law, amounts to (32.6).

To the existence of a constant of the motion there corresponds a simple property of the wave function. If the operator \hat{I} corresponds to a certain conserved quantity, then the Schrödinger equation will be satisfied not only by the wave function ψ but also by the wave function

$$\psi' = e^{i\alpha\hat{I}}\psi, \quad (32.7)$$

where α is an arbitrary real number. By definition

$$e^{i\alpha\hat{I}} = 1 + i\alpha\hat{I} + \frac{(i\alpha)^2}{2!}\hat{I}^2 + \dots$$

Substituting ψ' into the Schrödinger equation, we find

$$i\hbar \frac{\partial \psi'}{\partial t} = i\hbar \frac{\partial}{\partial t}(e^{i\alpha\hat{I}}\psi) = \hat{H}e^{i\alpha\hat{I}}\psi. \quad (32.8)$$

But, since \hat{I} , as the operator of a conserved quantity, satisfies the commutation condition $\hat{I}\hat{H} - \hat{H}\hat{I} = 0$, $\partial\hat{I}/\partial t = 0$, we have

$$\frac{\partial}{\partial t}(e^{i\alpha\hat{I}}\psi) = e^{i\alpha\hat{I}}\frac{\partial \psi}{\partial t}, \quad \hat{H}e^{i\alpha\hat{I}}\psi = e^{i\alpha\hat{I}}\hat{H}\psi,$$

and eq. (32.8) is satisfied directly.

Let us consider some simple examples. We begin with the case of a free particle. Then the Hamiltonian is of the form

$$\hat{H} = \frac{1}{2m}(\hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2).$$

Evidently, $[\hat{H}, \hat{p}_x] = [\hat{H}, \hat{p}_y] = [\hat{H}, \hat{p}_z] = 0$. Consequently,

$$\hat{p}_x = \hat{p}_y = \hat{p}_z = 0. \quad (32.9)$$

If at a certain initial instant the free particle was in a state with definite momentum, then this value of the momentum is conserved in time.

As another example let us consider a particle moving in the field produced by an infinite uniform plane (xy -plane). The potential energy of a particle in such a field depends only on the distance from the plane $U = U(|z|)$, so that the Hamiltonian is of the form

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla^2 + U(|z|).$$

The operators $\hat{p}_x, \hat{p}_y, \hat{l}_z$ commute with such a Hamiltonian. This means that in the case of motion in the field of a uniform plane (xy) the components of the momentum of the particle, p_x and p_y , and the z -component of the angular momentum, l_z , are conserved.

§33. Parity

The conservation laws considered above – the laws of conservation of energy, momentum and angular momentum – are the quantum-mechanical analogues of the conservation laws of classical mechanics. It turns out that in quantum mechanics there are also conservation laws which have no classical analogue. One such law is closely associated with the properties of space and is of a very general character. Namely, the Hamiltonian of a closed system must not change under the following transformations of the coordinates:

- (1) translation of the origin by an arbitrary segment;
- (2) rotation through an arbitrary angle;
- (3) inversion, i.e. the substitution $x_i \rightarrow -x_i$, in which the signs of all coordinates are changed.

As we have seen in the preceding section, the first two transformations are associated with the laws of conservation of momentum and angular momentum. In quantum mechanics it turns out that inversion is associated with still another general conservation law. As for the translation and rotation operators, which have been introduced earlier, one can also introduce the corresponding inversion operator \hat{I}

$$\hat{I}\psi(\mathbf{r}, t) = a\psi(-\mathbf{r}, t), \quad (33.1)$$

where a is a constant.

When the inversion operator \hat{I} is applied twice we arrive at the initial state. Hence it follows that $a^2 = 1$, i.e. $a = \pm 1$. Thus, in general, the following is

fulfilled:

$$\hat{I}\psi(\mathbf{r}, t) = \pm \psi(-\mathbf{r}, t), \quad (33.2)$$

i.e. the wave function itself, and not only the argument \mathbf{r} on which it depends, can change sign directly under inversion. Whether the transformation of the wave function under inversion will have $a = +1$ or $a = -1$ depends on the intrinsic properties of the particles described by this wave function.

The particles which are described by wave functions satisfying the condition

$$\hat{I}\psi(\mathbf{r}, t) = \psi(-\mathbf{r}, t)$$

are said to possess even intrinsic parity. On the contrary, particles which are described by wave functions satisfying the condition

$$\hat{I}\psi(\mathbf{r}, t) = -\psi(-\mathbf{r}, t)$$

have odd intrinsic parity.

We assume that the Hamiltonian of a closed system has the form

$$\hat{H} = \sum_i -\frac{\hbar^2}{2m_i} \nabla_i^2 + \frac{1}{2} \sum_{i \neq k} U_{ik}(|\mathbf{r}_i - \mathbf{r}_k|).$$

It is easily seen that this Hamiltonian does not change under the substitution $\mathbf{r}_i \rightarrow -\mathbf{r}_i$, i.e. it satisfies the condition $\hat{I}\hat{H}\psi = \hat{H}\hat{I}\psi$. This means that the operator \hat{I} commutes with the Hamiltonian

$$\hat{I}\hat{H} = \hat{H}\hat{I}. \quad (33.3)$$

We determine the eigenvalues λ of the inversion operator

$$\hat{I}\psi_\lambda(x) = \lambda\psi_\lambda(x). \quad (33.4)$$

We apply the inversion operator to this equation once more. Since under the two-fold inversion we come back to the initial value of the coordinates, this transformation is simply

$$\hat{I}^2\psi_\lambda = \psi_\lambda = \lambda\hat{I}\psi_\lambda = \lambda^2\psi_\lambda. \quad (33.5)$$

Whence we find that the eigenvalues λ are equal to ± 1 . A state with $\lambda = +1$ is said to have even parity or to be even. On the other hand, a state with $\lambda = -1$ has odd parity or is odd. If the parity operator commutes with the Hamiltonian operator, then the parity conservation law holds. The parity conservation law, like other conservation laws, imposes definite restrictions upon possible changes of the states of a system. Namely, if the system was in an

even state, then it will remain in such a state, and not pass over into an odd state. Naturally, the situation is analogous in the case of a system in an odd state.

Let us determine the parity of the state of a particle with angular momentum l . The fact that the angular momentum and parity can be determined simultaneously follows from the commutation of the corresponding operators:

$$\{\hat{I}, \hat{I}_x\} = 0; \quad \{\hat{I}, \hat{I}_y\} = 0; \quad \{\hat{I}, \hat{I}_z\} = 0; \quad \{\hat{I}, \hat{I}^2\} = 0. \quad (33.6)$$

From the expressions for the angular momentum operators $\hat{I}_x, \hat{I}_y, \hat{I}_z$ it is clear that they do not change under inversion. In the spherical system of coordinates the inversion has the form

$$r \rightarrow r; \quad \vartheta \rightarrow \pi - \vartheta; \quad \varphi \rightarrow \varphi + \pi. \quad (33.7)$$

The dependence of the wave function of a particle with definite angular momentum l on the angles ϑ, φ is given by the spherical function $Y_m(\vartheta, \varphi)$ (see §30). Under inversion (33.7) we have $\cos \vartheta \rightarrow -\cos \vartheta$ and $e^{im\varphi} \rightarrow (-1)^m e^{im\varphi}$. From formula (30.17) it is easily established how the associated Legendre polynomial $P_l^m(\xi)$ is transformed under a change of sign of its argument. Since $P_l(-\xi) = (-1)^l P_l(\xi)$, we obtain that $P_l^m(-\xi) = (-1)^{l+m} P_l^m(\xi)$. Taking into account the factor $(-1)^m$ which is given by the function $e^{im\varphi}$, we find that under inversion the wave function on the whole is multiplied by the factor $(-1)^l$. Taking into account also the factor $a = \pm 1$ associated with the intrinsic properties of particles, we get

$$\lambda = (-1)^l a. \quad (33.8)$$

Thus the states with even l have even parity if $a = 1$, and odd parity if $a = -1$. The states with odd l have, correspondingly, odd parity if $a = 1$, and even parity if $a = -1$. If we have a system of non-interacting particles, then the parity of the system is determined by the product of the parities of the individual particles. Indeed, in §14 we have seen that the wave function of a system of non-interacting particles can be written in the form of the product of the wave functions of the individual particles. Hence it follows immediately that under inversion the parities of individual particles are multiplied. If each of the particles is in a state with definite angular momentum (motion in a central field), then the parity of the entire system can be written in the form

$$\lambda = (-1)^{\sum k l_k} \prod_k a_k, \quad (33.9)$$

where the second factor is determined by the product of the intrinsic parities of the particles.

In addition to other conservation laws the parity conservation law is one of the most general laws of nature. The impossibility of transitions of a closed quantum-mechanical system from states with one parity into states with another parity — so-called forbidden transitions — is confirmed by a vast amount of experimental data in atomic as well as nuclear physics. However, it has been established (see §122) that the parity conservation law is not a universal physical law. The parity conservation law is violated in certain processes involving elementary particles.

§ 34. The uncertainty relation for time and energy

The relation between the uncertainty in the energy ΔE and a time interval Δt can be derived from the general apparatus of quantum mechanics, as was shown by Mandelshtam and Tamm*. Indeed, the total energy of a closed system can have no definite values which are constant in time. As we have explained in §32, its mean value and the probabilities of observing one or other possible value are constant in time. In other words, the form of the energy distribution function is conserved in time.

Knowing the distribution function one can define the value of the root-mean-square deviation of the energy ΔE , which is naturally also conserved in time in the usual way. The energy will have a definite value ($\Delta E = 0$) only if the system is in a stationary state. A characteristic indication of a stationary state is the constancy in time of the physical quantities of a given system.

Let us assume that a closed system is in a state with indefinite energy E at the initial instant of time. Further, let R be a quantity whose operator \hat{R} does not depend explicitly on time. For the given quantity one can, in the usual way, define its root-mean-square deviation ΔR and the mean value \bar{R} . Making use of (24.5) and (31.5), we write the relations

$$\Delta E \Delta R \geq \frac{1}{2} |(\overline{\hat{H}\hat{R}} - \hat{R}\hat{H})|, \quad (34.1)$$

$$\hbar \dot{\bar{R}} = i(\overline{\hat{H}\hat{R}} - \hat{R}\hat{H}). \quad (34.2)$$

* L.I.Mandelshtam and I.E.Tamm, *Izvestiya Akad. Nauk SSSR, physical series*, 9 (1945) 122.

Substituting (34.2) into (34.1), we obtain correspondingly

$$\Delta E \Delta R \geq \frac{1}{2} \hbar |\dot{\bar{R}}|. \quad (34.3)$$

This relation connects the uncertainty in the energy ΔE , the uncertainty ΔR in the value of R , and the rate of change of the mean value of the quantity R . Relation (34.3) can be rewritten in a somewhat more convenient form if one introduces the interval Δt — the time for which the mean value of R changes by an amount of the order of magnitude of its root-mean-square deviation ΔR

$$\Delta t = \frac{\Delta R}{\dot{\bar{R}}}. \quad (34.4)$$

Then we have

$$\Delta E \Delta t \geq \frac{1}{2} \hbar. \quad (34.5)$$

In particular, it follows from (34.3) that for the value of \bar{R} to change with time, R must possess a dispersion different from zero.

Thus we see that there is a definite relation between the dispersion of the total energy of the system and the rate of change of arbitrary quantities characterizing the system under consideration.

As a simple example let us consider a one-dimensional wave packet. We take the coordinate x as the quantity R , $R = x$. Then ΔR is the width of the packet, and Δt is the time of flight of the packet past a certain point of space. The relation (34.5) shows that the transit time essentially depends on the dispersion of the total energy ΔE .

From the inequality (34.5) then also follows a definite relation between the lifetime of a given state and the uncertainty in the energy, ΔE , of this state. Thus, assuming Δt to be equal to τ , the half-life, we obtain that in order of magnitude

$$\Gamma \sim \frac{\hbar}{2} \frac{1}{\tau}, \quad (34.6)$$

where Γ is the uncertainty in the energy of the initial state and gives the width of the corresponding spectral line. The problem of the connection of the decay law with the energy distribution function is considered in more detail in a work of Krylov and Fok*. In this work it is also shown that relation (34.5) cannot be applied to the measurement processes, because it is derived by making use of the Schrödinger equation. This follows, for example

* I.S.Krylov and V.A.Fok, J. Exp. Theor. Phys. (USSR), 17 (1947) 93.

from the fact that a given object during the process of measurement is no longer a closed quantum-mechanical system.

For the measurement processes the corresponding inequality must be formulated in the form of a certain physical principle (stated by Bohr)

$$\Delta(E - E') \Delta t > \hbar, \quad (34.7)$$

where E and E' are the values of the energy of the object before and after the measurement process, and $\Delta(E - E')$ is the absolute value of the uncertainty in the measurement of the energy of the object, i.e. the corresponding error of the measurement if it was carried out during a time Δt .

Relation (34.7) is very important in the analysis of the results of measurements, i.e. for an experimental check of the results given by quantum mechanics. We shall illustrate it by the simple example of a free particle. For the measurement of the quantities E, p, v (energy, momentum, velocity) of the particle it is necessary to consider the collision of this particle with another system (apparatus). Assuming for simplicity that the motion is one-dimensional, we write the momentum conservation law

$$p + k - p' - k' = 0. \quad (34.8)$$

Here we denote by k and k' the momentum of the apparatus before and after the collision. We shall denote by primes the quantities referring to the systems after the collision. It can be assumed that the momentum of the apparatus before and after the collision is accurately measured. Then from relation (34.8) there follows the equality of the errors in the measurement of the momentum of the particle before and after the collision

$$\Delta p = \Delta p'. \quad (34.9)$$

The error in the measurement of the energy can be expressed in terms of the error in the measurement of the momentum, since

$$\Delta E = \frac{\partial E}{\partial p} \Delta p = v \Delta p,$$

$$\Delta E' = \frac{\partial E'}{\partial p'}, \Delta p' = v' \Delta p'.$$

In view of the equality of the errors Δp and $\Delta p'$, we have

$$\Delta(E - E') = |v - v'| \Delta p. \quad (34.10)$$

We multiply (34.10) by the time of measurement Δt . We then obtain

$$\Delta(E - E') \Delta t = |v - v'| \Delta p \Delta t. \quad (34.11)$$

But the quantity $|v - v'| \Delta t$ represents an additional error in the coordinate, which appeared during the time of measurement Δt . The total uncertainty in the coordinate Δx can be written in the form

$$\Delta x = (\Delta x)_0 + |v - v'| \Delta t,$$

where $(\Delta x)_0$ is the uncertainty in the coordinate of the particle which existed before the collision being considered. In particular, $(\Delta x)_0$ can be made arbitrarily small. The fact that the value of $(\Delta p)_0$ will then be large is of no importance, because $(\Delta p)_0$ is in no way connected with the error Δp considered.

Heisenberg's uncertainty relation $\Delta p \Delta x \gtrsim \hbar$ must be fulfilled irrespective of the value of $(\Delta x)_0$. Consequently

$$|v - v'| \Delta t \Delta p > \hbar. \quad (34.12)$$

Comparing with (34.11), we arrive at the inequality

$$\Delta(E - E') \Delta t > \hbar \quad (34.13)$$

in accordance with (34.7). We see that the error in the measurement of the energy tends to zero provided the measurement process lasts a sufficiently long time (in the limit $\Delta t \rightarrow \infty$).

We note in addition that, as follows from (34.12), the measurement of the momentum for a given value of the error Δp leads to a change in the velocity of the particle,

$$|v - v'| > \frac{\hbar}{\Delta p \Delta t},$$

and, consequently, to a change in the momentum. Only if the measurement is carried out during an infinitely long time ($\Delta t \rightarrow \infty$) does the momentum not change. Of course, a measurement of the momentum over a long period of time can make sense only if the particle is free. Thus we see that the process of measurement of the momentum in small time intervals is irreproducible. The measurement brings the micro-object into a completely new state (see §5).

Motion in a Centrally Symmetric Field

§35. The Schrödinger equation

We can now apply the mathematical apparatus of quantum mechanics, developed in the preceding chapter to the study of the properties of real systems. It is natural to consider, first of all, the hydrogen atom, the simplest atomic system. In the hydrogen atom the potential energy of interaction of the electron with the nucleus depends only on the distance between them, $|\mathbf{r}_1 - \mathbf{r}_2|$. The problem of the motion of two particles with the interaction law $U(|\mathbf{r}_1 - \mathbf{r}_2|)$ amounts, as we have explained in §14, to the problem of the motion of one particle with reduced mass μ in a field $U(r)$. In view of the large difference in the masses, the reduced mass μ is very close to the mass of the electron. If also the size of the proton is neglected, then the hydrogen atom represents an electron moving in the Coulomb field of a motionless centre. Such a field is a particular case of a centrally symmetric field in which the potential energy depends only on the distance from the force centre. We shall first consider the motion of an electron in a centrally symmetric field of the most general form, after which we shall pass over to the case of the Coulomb field.

The Schrödinger equation for the stationary states of a particle moving in

a force field with potential energy $U(r)$ has the form

$$\nabla^2 \psi + \frac{2\mu}{\hbar^2} [E - U(r)] \psi = 0. \quad (35.1)$$

In the case of a centrally symmetric potential field it is convenient to transform the Schrödinger equation to spherical coordinates, since the potential energy depends only on the distance from the origin r . Expressing the Laplacian operator in spherical coordinates, we have

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2} \nabla_{\vartheta\varphi}^2 \psi + \frac{2\mu}{\hbar^2} [E - U(r)] \psi = 0. \quad (35.2)$$

This equation is conveniently transformed by introducing into it explicitly the operator of the square of the angular momentum $\hat{\mathbf{I}}^2$. Substituting its value according to formula (30.9), we have

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) - \frac{\hat{\mathbf{I}}^2}{\hbar^2 r^2} \psi + \frac{2\mu}{\hbar^2} [E - U(r)] \psi = 0. \quad (35.3)$$

We shall first of all show that in the case of motion in a centrally symmetric field two more conservation laws are satisfied, in addition to the energy conservation law: the total angular momentum conservation law and the law of conservation of the z -component of the angular momentum where the z -axis is arbitrarily oriented in space. When we speak here of the conservation of total angular momentum we mean the quantity described by the operator $\hat{\mathbf{I}}^2$ (the square of the angular momentum). For this, according to general rules, we consider the conditions for the commutation of the operators $\hat{\mathbf{I}}^2$ and \hat{I}_z with the Hamiltonian. It is obvious that in our case the Hamiltonian \hat{H} can be written in the form

$$\hat{H} = -\frac{\hbar^2}{2\mu} \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{\hat{\mathbf{I}}^2}{2\mu r^2} + U(r). \quad (35.4)$$

The operator $\hat{\mathbf{I}}^2$ involves only the angular variables ϑ, φ , and the differential operators with respect to these variables. Hence the operator $\hat{\mathbf{I}}^2$ commutes with any operator of differentiation with respect to r , as well as with the operator of the coordinate r itself

$$\hat{H}\hat{\mathbf{I}}^2 - \hat{\mathbf{I}}^2\hat{H} = 0. \quad (35.5)$$

An analogous relation also holds for the operator \hat{I}_z in view of the fact that, as we have seen in §30, it commutes with the operator $\hat{\mathbf{I}}^2$ (30.4):

$$\hat{H}\hat{I}_z - \hat{I}_z\hat{H} = 0. \quad (35.6)$$

Since, in motion in a centrally symmetric field, three quantities are conserved — the energy, the square of the angular momentum \mathbf{l}^2 , and the projection l_z of the angular momentum onto an arbitrary axis — we shall consider states with given values of the three quantities.

It should be noted that in motion in a centrally symmetric field the laws of conservation of energy, of total angular momentum and of the z -component of angular momentum also hold in classical mechanics.

We have considered previously the states of a system with given values of the total angular momentum and its projection onto the z -axis. The eigenvalues of the operators $\hat{\mathbf{l}}^2$ and \hat{l}_z were characterized by the azimuthal and magnetic quantum numbers l and m , while the spherical functions $Y_{lm}(\vartheta, \varphi)$ with the indices l, m were the eigenfunctions of these operators.

Equation (35.3) allows one to separate the variables. Its angular part is the same as eq. (30.14). It describes the motion with given values l and m . Hence it is natural to seek the solution of (35.3) in the form

$$\psi(r, \vartheta, \varphi) = R(r) Y_{lm}(\vartheta, \varphi). \quad (35.7)$$

Substituting expression (35.7) into eq. (35.3) and taking into account that $\hat{\mathbf{l}}^2 Y_{lm} = \hbar^2 l(l+1) Y_{lm}$, we arrive at the following equation for the radial part of the wave function $R(r)$:

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{2\mu}{\hbar^2} \left(E - U(r) - \frac{\hbar^2}{2\mu} \frac{l(l+1)}{r^2} \right) R = 0. \quad (35.8)$$

We see that the expression for the radial component R of the wave function ψ depends essentially on the form of the potential energy $U(r)$. At the same time the angular part $Y_{lm}(\vartheta, \varphi)$ of the wave function is determined only by the value of the angular momentum of the particle (the number l) and its z -component (the number m). States with a given angular momentum are denoted by small letters:

$$\begin{array}{cccccccc} l = & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ & s & p & d & f & g & h & i & k \end{array}$$

Also the parity of the state is determined by the value of the quantum number l . In §33 we have shown that in a state with given total angular momentum and given z -component of angular momentum, the parity is equal to $(-1)^l$, i.e. under inversion the spherical function Y_{lm} goes over into $(-1)^l Y_{lm}$. Since the radial wave function, which depends on the absolute value of the radius vector, does not change under inversion, the transforma-

tion law mentioned also refers to the total wave function

$$\psi(r, \vartheta, \varphi) \rightarrow (-1)^l \psi(r, \vartheta, \varphi).$$

Thus the states s, d, g, ... are even, while p, f, h, ... are odd (for even intrinsic parity).

The probability that an electron in the state $\psi(r, \vartheta, \varphi) = R(r) Y_{lm}(\vartheta, \varphi)$ will be observed in an infinitesimal volume element with coordinates r, ϑ, φ , is given by the formula

$$dW(r, \vartheta, \varphi) = |\psi(r, \vartheta, \varphi)|^2 r^2 dr d\Omega, \quad (35.9)$$

where $d\Omega = \sin \vartheta d\vartheta d\varphi$. If this expression is integrated with respect to all values of the angles ϑ, φ , then we shall obtain the probability of observing the electron in a spherical layer between r and $r+dr$

$$dW(r) = |R(r)|^2 r^2 dr. \quad (35.10)$$

Integrating (35.9) with respect to all values of the radius r from 0 to ∞ , we find the probability $dW(\vartheta, \varphi)$ of observing the electron in the solid angle $d\Omega$ in the direction defined by the angles ϑ, φ

$$dW_{lm}(\vartheta, \varphi) = |Y_{lm}|^2 d\Omega. \quad (35.11)$$

It follows from the definition of the spherical function (30.16) that the last expression does not depend on the angle φ . This means that in the plane perpendicular to the z -axis the distribution of the probability of finding the particle is completely symmetric. It should be noted that we understand the z -axis to be an arbitrarily chosen direction in space; the projection of the angular momentum onto this direction is fixed. Thus it follows from (35.11) that

$$dW_{lm} \sim |P_l^m(\cos \theta)|^2 d\Omega. \quad (35.12)$$

The probability distribution (35.12) is determined by the two quantum numbers l and m , i.e. it depends on the value of the total angular momentum and its projection on the z -axis.

The state with $l=0$ (s-state) possesses spherical symmetry, because for $l=0$ (consequently also $m=0$) $P_0^0 = \text{const}$

$$dW_{00} = \frac{1}{4\pi} d\Omega. \quad (35.13)$$

In the p-state ($l=1$) the probability distribution is given by the following

expressions:

$$dW_{1,\pm 1} = \frac{3}{8\pi} \sin^2 \vartheta d\Omega, \quad dW_{10} = \frac{3}{4\pi} \cos^2 \vartheta d\Omega. \quad (35.14)$$

The distributions (35.12) for different l and m are presented graphically in fig. V.9 in the form of polar diagrams. The probability $dW_{lm}/d\Omega$ is plotted on the radius vector drawn at the angle ϑ to the z -axis.

Let us consider eq. (35.8) for the radial component of the wave function in more detail. First of all, it follows from this equation that the energy of the particle does not depend on the z -component of the angular momentum. This is associated with the fact that in a spherically symmetric field all directions are equivalent. Thus the isotropy of space leads to degeneracy of the levels of the system in which the energy does not depend on the quantum number m . It should be noted that degeneracy is always due to definite symmetry properties of the system considered.

Instead of the function R it is convenient to introduce the function $\chi(r)$:

$$R(r) = \frac{1}{r} \chi(r). \quad (35.15)$$

For $\chi(r)$ we find

$$\frac{d^2\chi}{dr^2} + \frac{2\mu}{\hbar^2} \left(E - U(r) - \frac{\hbar^2}{2\mu} \frac{l(l+1)}{r^2} \right) \chi = 0. \quad (35.16)$$

The condition of the finiteness of the wave function for $r = 0$ leads to the requirement

$$\chi(0) = 0. \quad (35.17)$$

The equation for the radial function (35.8) amounts to the equation of a one-dimensional motion with an effective potential energy equal to

$$U_{\text{eff}}(r) = U(r) + \frac{\hbar^2}{2\mu} \frac{l(l+1)}{r^2}. \quad (35.18)$$

As in classical mechanics, the quantity $\hbar^2 l(l+1)/2\mu r^2$ is called the centrifugal energy.

Without fixing the detailed form of the potential energy $U(r)$, one can nevertheless make definite conclusions about the behaviour of the wave function near the origin and at very large distances from the force centre.

Let us first study the region of small distances $r \rightarrow 0$. We assume that near the origin the potential energy of interaction $U(r)$ changes so slowly that the

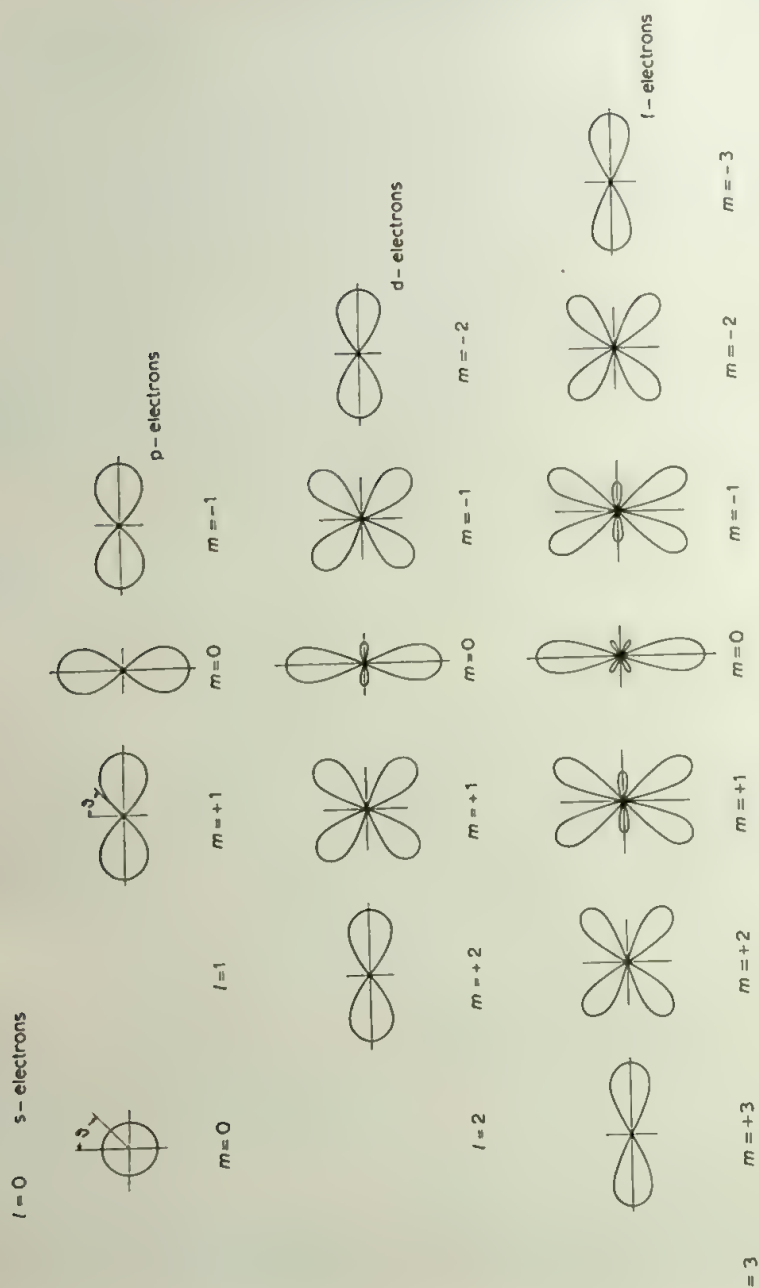


Fig. V.9

following condition holds:

$$\lim_{r \rightarrow 0} r^2 U(r) = 0. \quad (35.19)$$

This condition means that $|U(r)|$ for $r \rightarrow 0$ increases more slowly than $1/r^2$. It is fulfilled, in particular, for the electron in the Coulomb field of the nucleus. Then in eq. (35.16) for $r \rightarrow 0$ the terms $E\chi$ and $U(r)\chi$ can be disregarded in comparison with the term $\hbar^2 l(l+1)/2\mu r^2 \chi$, and we obtain

$$\frac{d^2 \chi}{dr^2} - \frac{l(l+1)}{r^2} \chi = 0.$$

We seek the solution of the last equation in the form $\chi = Ar^\gamma$. Substituting this expression into the equation, we have

$$\gamma(\gamma-1) = l(l+1). \quad (35.20)$$

Equation (35.20) has two roots: $\gamma_1 = l+1$; $\gamma_2 = -l$. We must discard the second root, since it corresponds to a function R which increases indefinitely for $r \rightarrow 0$. Thus we find that at small distances $\chi(r) \sim r^{l+1}$, and the radial part of the wave function is expressed by the formula

$$R(r) = Ar^l. \quad (35.21)$$

The probability of finding the particle at a given distance r from the centre, independent of the angles ϑ and φ , is given by the square of the modulus of the radial function, i.e. by the quantity $|R|^2 r^2 dr$.

It follows from (35.21) that for small r this probability is proportional to $r^{2l+2} dr$ and is smaller the larger l becomes. The centrifugal force acts as if to throw the particle out from the centre.

Further, we study the asymptotic behaviour of the wave function at large distances from the origin. At such distances the force acting on the particle tends to zero, and consequently, the potential energy $U(r)$ tends to a constant. If not specified otherwise, we shall choose this constant as the zero of potential energy, i.e. we shall assume that $\lim_{r \rightarrow \infty} U(r) = 0$. Then in eq. (35.16), for large r , the terms $U\chi$ and $\hbar^2 l(l+1)/2\mu r^2 \chi$ can be disregarded* in comparison with the term $E\chi$. In this case eq. (35.16) assumes the

* From a more detailed analysis it follows that this is legitimate if the potential energy at infinity decreases according to the law $1/r^n$, where $n > 1$. See, for example, L.D. Landau and E.M. Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965); V.A. Fok, *Nachala kvantovoi mekhaniki (Principles of quantum mechanics)* (KUBUCH, 1932) p. 126.

form

$$\frac{d^2\chi}{dr^2} + k^2\chi = 0, \quad k = \left(\frac{2\mu E}{\hbar^2}\right)^{\frac{1}{2}}. \quad (35.22)$$

The solution of this last equation can obviously be written as

$$\chi = A_1 e^{ikr} + A_2 e^{-ikr}, \quad (35.23)$$

where A_1 and A_2 are integration constants.

Let us consider, first of all, the solutions corresponding to positive values of the energy. For $E > 0$ the quantity k defined by formula (35.22) is real. The radial part of the wave function (35.15) amounts to the sum of the two functions

$$R(r) = A_1 \frac{e^{ikr}}{r} + A_2 \frac{e^{-ikr}}{r}. \quad (35.24)$$

Since both terms are restricted in modulus, neither of the constants A_1 and A_2 can be equal to zero. At a large distance from the force centre the radial function represents the superposition of a converging and a diverging spherical wave.

A definite conclusion can be also made about the energy spectrum of a particle for an arbitrary form of the energy of interaction $U(r)$. Indeed, the function (35.23) does not reduce to zero at infinity, which corresponds to an infinite motion, i.e. a motion in which the particle or the system goes off to infinity. The integral of the square of the modulus of function (35.24), taken over all space, diverges. But, as we have noted in § 16, such functions correspond to a continuous spectrum. Consequently, for $E > 0$ the energy spectrum is continuous. If the radial component of the current density is equal to zero, then function (35.24) must be real. Correspondingly we assume that

$$A_1 = \frac{1}{2i} A' e^{i\alpha}, \quad A_2 = -\frac{1}{2i} A' e^{-i\alpha}, \quad (35.25)$$

A' and α being real.

Then corresponding to (35.24) the radial function R assumes the form

$$R = A' \frac{\sin(kr + \alpha)}{r}, \quad (35.26)$$

where the phase α depends on k , l , as well as on the actual form of the function $U(r)$. In the following section we shall show that for a free particle

$$(U \equiv 0)$$

$$\alpha = -\frac{1}{2}l\pi.$$

In accordance with this we assume that

$$\alpha = -\frac{1}{2}l\pi + \delta_l, \quad (35.27)$$

where the phases δ_l are directly connected with the action of the force field on the particle and reduce to zero for free motion.

We now consider the region of negative energies, $E < 0$. Since the kinetic energy of the particle is always positive, the total energy can be negative only in the case of attraction of the particle towards the centre. If $E < 0$, the quantity k has purely imaginary values, i.e. $k = i\kappa$, where $\kappa = (-2\mu E/\hbar^2)^{\frac{1}{2}}$. The radial function (35.24) is written in the form

$$R = A_1 \frac{e^{-\kappa r}}{r} + A_2 \frac{e^{\kappa r}}{r}. \quad (35.28)$$

In order to satisfy the requirement of the finiteness of the wave function for $r \rightarrow \infty$, we have to assume that the constant A_2 is equal to zero

$$R = A_1 \frac{e^{-\kappa r}}{r}. \quad (35.29)$$

Then the radial wave function R tends to zero as $r \rightarrow \infty$. This means that the probability of finding the particle at an infinitely large distance from the force centre is equal to zero. Consequently the motion of the particle is finite. We see that there is a similarity between the conclusions of quantum and classical mechanics: for a positive total energy ($E > U(\infty)$) the particles go off to infinity, while for a negative total energy they perform a finite motion.

Let us now consider the energy spectrum for $E < 0$. As we have explained, a finite motion corresponds to these energies and the corresponding wave functions (35.29) are quadratically integrable. Such wave functions, as was pointed out in §16, belong to a discrete spectrum. Consequently, for $E < 0$ we have a discrete energy spectrum.

The general solution of the Schrödinger equation (35.2) can be written in the form of a superposition of the wave functions (35.7)

$$\psi(r, \vartheta, \varphi) = \sum_{l,m} B_{lm} R_l(r) Y_{lm}(\vartheta, \varphi). \quad (35.30)$$

For a solution which does not depend on the angle φ we obtain a simpler

expression (superposition of states with $m = 0$)

$$\psi(r, \vartheta) = \sum_l c_l R_l(r) P_l(\cos \vartheta). \quad (35.31)$$

§36. The free motion of a particle with given angular momentum

So far we have represented a freely moving particle by a plane wave $e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}$, where \mathbf{k} is the wave vector of the particle $\mathbf{k} = \mathbf{p}/\hbar$, and $\omega = E/\hbar$. This wave function describes a stationary state with a definite value of the momentum and energy $E = p^2/2m$ of the particle. For what follows, we need to find the wave functions of the stationary states of a freely moving particle, in which, in addition to a definite value of the energy E , the values of the angular momentum and the z -component of the angular momentum are also given. In classical mechanics a free particle moving with a definite momentum also possesses automatically a definite angular momentum. In quantum mechanics the situation is fundamentally altered. In a state with a given momentum the angular momentum is an indefinite quantity. On the other hand, in a state where the angular momentum and its projection onto the z -axis are given the direction of the momentum is indefinite. This is associated with the fact that the corresponding quantities cannot simultaneously have sharp values.

In order to find the wave function required, let us consider the motion of a free particle in spherical coordinates. Setting $U(r) \equiv 0$ in the Schrödinger equation (35.3), we have

$$-\frac{\hbar^2}{2m} \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \frac{\hat{L}^2}{2mr^2} \psi = E\psi. \quad (36.1)$$

We seek the wave function of the free particle in the form

$$\psi_{klm}(r, \vartheta, \varphi) = R_{kl}(r) Y_{lm}(\vartheta, \varphi). \quad (36.2)$$

In this case the radial function R_{kl} must satisfy eq. (35.8) in which one must set $U \equiv 0$

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dR_{kl}}{dr} \right) + \left(k^2 - \frac{l(l+1)}{r^2} \right) R_{kl} = 0. \quad (36.3)$$

Here we have expressed the energy E in terms of the wave number k . For

$l = 0$ the equation is rewritten in the form

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dR_{k0}}{dr} \right) + k^2 R_{k0} = 0. \quad (36.4)$$

The solution of the above equation which does not go to infinity at the origin is the function

$$R_{k0} = A \frac{\sin kr}{r}. \quad (36.5)$$

To find the solution of eq. (36.3) for $l \neq 0$ we introduce a new function given by the formula

$$R_{kl} = s^{-\frac{1}{2}} Z \quad (36.6)$$

where $s = kr$. For such a substitution eq. (36.3) is easily transformed to the form

$$\frac{d^2 Z}{ds^2} + \frac{1}{s} \frac{dZ}{ds} + \left(1 - \frac{(l + \frac{1}{2})^2}{s^2} \right) Z = 0. \quad (36.7)$$

The solution of eq. (36.7) satisfying the condition of finiteness of the wave function at the origin is a Bessel function of half-integer order

$$Z(s) = C J_{l+\frac{1}{2}}(s). \quad (36.8)$$

Correspondingly for the radial function we have

$$R_{kl} = (kr)^{-\frac{1}{2}} C J_{l+\frac{1}{2}}(kr). \quad (36.9)$$

At a large distance from the origin ($r \rightarrow \infty$) one can make use of the known asymptotic expression for the Bessel function and obtain the asymptotic value $R_{kl}(r)$

$$R_{kl}(r) = C(2/\pi)^{\frac{1}{2}} \frac{\sin(kr - \frac{1}{2}l\pi)}{kr}. \quad (36.10)$$

The constant C is determined by the normalization condition. At small distances from the force centre ($r \rightarrow 0$) the radial function (36.9) assumes the form

$$R_{kl} \sim r^l \quad (36.11)$$

in accordance with the general expression (35.21).

§37. The spherical well

As a simple and at the same time important example we shall consider the motion of a particle in a centrally symmetric field defined by the expression

$$U(r) = \begin{cases} -U_0 & (r \leq a), \\ 0 & (r > a). \end{cases}$$

A field of this type is called a spherically symmetric potential well. The potential well shown in fig. V.10 represents an idealized model of a system in which the interaction with the centre is realized by so-called short-range forces. Short-range forces are understood to be forces which decrease with distance so rapidly that they can be assumed to be practically equal to zero at distances exceeding a certain distance, a , called the range of the short-range force. The importance of the consideration of systems with short-range forces is clear, for example, from the fact that the forces of interaction between nucleons, nuclear forces, are of such a type.

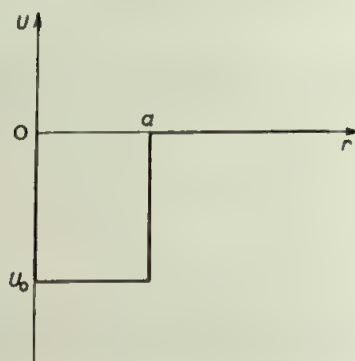


Fig. V.10

The idealization of a system by means of the model of a spherical potential well amounts to the assumptions of total isotropy of the forces and the constancy of the potential energy for $r < a$.

For simplicity let us consider the motion of a particle with angular momentum $l = 0$. It is obvious that two different modes of motion are possible. For $E < 0$ the total energy of the particle is smaller than the potential energy at infinity, which corresponds to a finite motion. On the contrary, for $E > 0$ there is an infinite motion. To the first case, to which we now confine

ourselves, there corresponds a discrete energy spectrum, while to the second case there corresponds a continuous spectrum.

The wave function of a particle with $l = 0$ depends only on the coordinate r , and not on the angles ϑ and φ . Upon the substitution $\chi(r) = rR(r)$ the Schrödinger equation will have the form (35.16)

$$\frac{d^2\chi}{dr^2} + \frac{2m}{\hbar^2} (E + U_0)\chi = 0 \quad (r \leq a), \quad (37.1)$$

$$\frac{d^2\chi}{dr^2} + \frac{2m}{\hbar^2} E\chi = 0 \quad (r > a). \quad (37.2)$$

We write the solution of eq. (37.1) in the form

$$\chi(r) = A \sin \kappa r + B \cos \kappa r, \quad (37.3)$$

where

$$\kappa = \left(\frac{2m}{\hbar^2} (U_0 - |E|) \right)^{\frac{1}{2}}.$$

For the wave function R to be finite at the origin it is necessary to set $\chi(0) = 0$. Consequently, inside the well the solution of eq. (37.1) has the form

$$\chi(r) = A \sin \kappa r. \quad (37.4)$$

The solution outside the well which reduces to zero at infinity is expressed by the formula

$$\chi(r) = B e^{-\kappa' r}, \quad (37.5)$$

where κ' denotes the quantity $\kappa' = [(2m/\hbar^2)|E|]^{\frac{1}{2}}$.

It follows from the continuity of the wave function that solution (37.4) must go over continuously into solution (37.5) at the surface of the sphere $r = a$. The derivative of the wave function also must be continuous at this surface. Hence we can equate to each other the logarithmic derivative of the functions (37.4) and (37.5) for $r = a$. We then obtain

$$\kappa \cotan \kappa a = -\kappa'. \quad (37.6)$$

This relation can be rewritten in the form

$$\sin \kappa a = \pm \left[\frac{\kappa'^2}{\kappa^2} + 1 \right]^{-\frac{1}{2}} \quad (37.7)$$

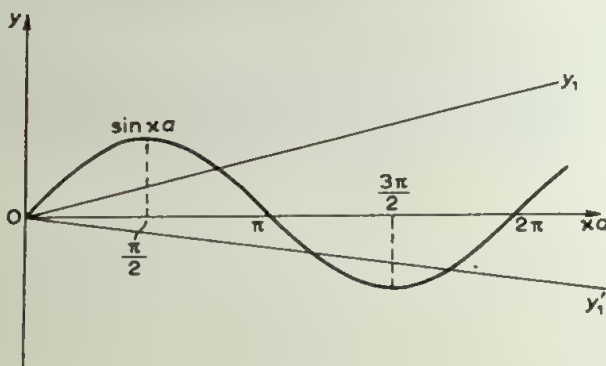


Fig. V.11

or, taking into account the expressions for κ and κ' , we have

$$\sin \kappa a = \pm \left(\frac{\hbar^2}{2mU_0 a^2} \right)^{\frac{1}{2}} \kappa a. \quad (37.8)$$

The roots of eq. (37.8) determine the energy levels of the particle in the well. Equation (37.8) is conveniently solved graphically. Namely, the roots of eq. (37.8) are the intersections of the straight lines

$$y_1 = \left(\frac{\hbar^2}{2mU_0 a^2} \right)^{\frac{1}{2}} \kappa a \quad \text{and} \quad y_1' = - \left(\frac{\hbar^2}{2mU_0 a^2} \right)^{\frac{1}{2}} \kappa a$$

with the curve $\sin \kappa a$ (see fig. V.11). Only crossover points for which $\cotan \kappa a$ has negative values may be chosen in correspondence with (37.6). From the graph in fig. V.11 it is seen that the roots of eq. (37.8) do not always exist. In order that a bound state (energy level) may exist the well must be sufficiently deep. Let us determine the minimum depth $U_{0\min}$ corresponding to the appearance of the first energy level. As is seen from fig. V.11, the first level will appear when the straight line passes through the peak of the sinusoidal curve at $\kappa a = \frac{1}{2}\pi$. The tangent of the slope angle is equal to $2/\pi$. Consequently, the minimum potential energy $U_{0\min}$ for which there is a bound state of the particle in the spherical well is determined by the condition

$$\left(\frac{\hbar^2}{2mU_{0\min} a^2} \right)^{\frac{1}{2}} = \frac{2}{\pi},$$

whence

$$U_{0\min} = \frac{\pi^2 \hbar^2}{8ma^2}. \quad (37.9)$$

We find the first energy level in the potential well of minimum depth $U_{0\min}$ from the condition $\kappa a = \frac{1}{2}\pi$ or

$$\left(\frac{2ma^2}{\hbar^2} (U_{0\min} - |E_1|) \right)^{\frac{1}{2}} = \frac{\pi}{2}.$$

Taking into account the value of $U_{0\min}$ we find that $E_1 = 0$, i.e. the energy of the particle in the first level is equal to zero and there are no other levels in the well. Also, the energy of the first level decreases with increasing depth of the well and becomes negative. In the graph this corresponds to a decrease in the slope of the straight line with respect to the abscissa. For a certain slope another root will appear in addition to the root corresponding to the first level. This new root corresponds to the appearance of a second energy level in the well. The number of crossover points in the graph increases with increasing depth of the well, which corresponds to an increase in the number of allowed energy levels of the particle in the potential well.

In conclusion we stress that the absence of bound states for a particle in a potential well of depth $U_0 < U_{0\min}$ represents a specific quantum-mechanical effect which has no analogue in classical physics. Indeed, however small the depth of the well in classical physics, a particle which falls into it with an initial kinetic energy less than the depth of the well will be confined in it. In quantum mechanics this proposition, in general, does not hold.

§38. Motion in a Coulomb field

As we have already pointed out, the most important example of the motion of a particle in a centrally symmetric field is the motion of an electron in the Coulomb field of the atomic nucleus. The simplest atomic system of such a kind, consisting of a nucleus and an electron, is the hydrogen atom, and also the ion of any atom in which only one electron remains. The mesohydrogen atom, consisting of a proton and a negatively charged meson, is another example.

The problem of the motion of two bodies, a nucleus and an electron, reduces to the problem of the motion of one particle with reduced mass μ in the Coulomb field (see §14).

It is clear that the theory of the hydrogen atom and hydrogen-like systems is extremely important, since these systems are the simplest atomic systems. Furthermore it turns out that in the case of the motion of a particle in the Coulomb field of a nucleus one can obtain a complete analytical solution of

the Schrödinger equation. This makes it possible to follow the appearance of general quantum-mechanical regularities in atomic systems.

The potential energy of an electron moving in the field of a nucleus with charge Ze is given by the formula

$$U(r) = -\frac{Ze^2}{r}. \quad (38.1)$$

We write the Schrödinger equation for the radial wave function (35.8)

$$\frac{d^2R}{dr^2} + \frac{2}{r} \frac{dR}{dr} - \frac{l(l+1)}{r^2} R + \frac{2\mu}{\hbar^2} \left(E + \frac{Ze^2}{r} \right) R = 0. \quad (38.2)$$

We are at first interested in states belonging to a discrete energy spectrum. In correspondence with § 16 these states correspond to a finite motion of the electron and, consequently, their energy is negative, $E < 0$ (see § 35).

In solving eq. (38.2) it is convenient to use dimensionless quantities. This will make all formulae less cumbersome.

We choose as basic quantities the charge of the electron e , its reduced mass μ , and the Planck constant \hbar . From these quantities one can make a combination having the dimensionality of a length

$$a = \frac{\hbar^2}{\mu e^2}. \quad (38.3)$$

As we shall see below, this length is a characteristic atomic dimension. If the reduced mass μ is set equal to the mass of the electron m , then $a = 0.529 \times 10^{-8}$ cm.

The system of units based on the quantities e , μ and a is called the Coulomb system.

The quantity e^2/\hbar , equal to $1/137$ of the velocity of light, will be the unit of velocity, while the quantity

$$E_0 = \frac{\mu e^4}{\hbar^2} = \frac{e^2}{a} \quad (38.4)$$

will be the unit of energy. For $\mu = m$, $E_0 = 4.30 \times 10^{-11}$ erg = 27.07 eV. We introduce into eq. (38.2) the dimensionless variable ρ and the energy ϵ

$$\rho = r/a, \quad \epsilon = -E/E_0. \quad (38.5)$$

Then this equation is rewritten in the form

$$\frac{d^2R}{d\rho^2} + \frac{2}{\rho} \frac{dR}{d\rho} + \left(-2\epsilon - \frac{l(l+1)}{\rho^2} + \frac{2Z}{\rho} \right) R = 0. \quad (38.6)$$

At small distances the function R behaves, according to (35.21), as ρ^l . At large distances this function has the form $R \sim \exp [-(2\epsilon)^{\frac{1}{2}}\rho]$ (see (35.29)). Corresponding to this we shall seek the solution of eq. (38.6) in the form

$$R(\rho) = \rho^l e^{-\beta\rho} v(\rho). \quad (38.7)$$

where $\beta = (2\epsilon)^{\frac{1}{2}}$. Substituting expression (38.7) into (38.6), we obtain after simple calculations

$$\rho \frac{d^2 v}{d\rho^2} + 2 \frac{dv}{d\rho} (l - \beta\rho + 1) + 2v(Z - \beta - \beta l) = 0. \quad (38.8)$$

We introduce a new variable

$$\xi = 2\beta\rho. \quad (38.9)$$

Denoting differentiation with respect to this new variable by a prime, we have

$$\xi v'' + v'(2l + 2 - \xi) + v\left(\frac{Z}{\beta} - l - 1\right) = 0. \quad (38.10)$$

The radial wave function R must remain finite over the entire region of variation of the variable ξ , for $\xi \rightarrow \infty$ as well as for $\xi \rightarrow 0$.

We seek the solution of eq. (38.10) in the form of a series

$$v(\xi) = \sum_{k=0}^{\infty} a_k \xi^k. \quad (38.11)$$

Substituting (38.11) into eq. (38.10) and gathering terms with the same powers of ξ , we obtain

$$\sum_k \xi^k \left[(k+1)(2l+2+k)a_{k+1} + \left(\frac{Z}{\beta} - l - 1 - k \right) a_k \right] = 0. \quad (38.12)$$

Equation (38.12) will be satisfied for arbitrary values of ξ if the coefficients of all powers of ξ are equal to zero. Hence, equating the square bracket to zero we arrive at the following recurrence formula:

$$a_{k+1} = \frac{k+l+1-(Z/\beta)}{(k+1)(2l+2+k)} a_k. \quad (38.13)$$

We note that the function v defined by the series (38.11), with coefficients a_k which satisfy (38.13), can be expressed in terms of the confluent hyper-

geometric function*

$$v = AF \left(1 + l - \frac{Z}{\beta}, 2l + 2, \xi \right). \quad (38.14)$$

It is easily shown, by analogy with what was done in § 10, that the series (38.11) diverges as e^k for $\xi \rightarrow \infty$. This means that if the wave function were expressed by the series (38.11) it would not satisfy the condition of being finite at arbitrarily large distances from the force centre. In order to define the function which possesses the necessary properties and is a solution of eq. (38.10) we must, as was done in solving the problem of the oscillator, cut off the series at a certain term, i.e. reduce it to a polynomial. If for a certain value of the number $k = n_r$ the coefficient a_{n_r+1} reduces to zero, then according to (38.13) all subsequent coefficients a_{n_r+2} , a_{n_r+3} and so on also reduce to zero. In this case the infinite series reduces to a polynomial of the n_r th degree. For large values of ξ the function $v(\xi)$ will increase according to the power law $v(\xi) \sim \xi^{n_r}$, while the wave function will tend to zero at infinity on account of the exponential factor. For $\xi \rightarrow 0$ the polynomial $v(\xi)$ tends to the constant quantity a_0 and the wave function (38.7) correspondingly reduces to zero or tends to a constant. Thus we see that the wave function will satisfy the standard boundary conditions.

Let us now consider the conditions under which the coefficient of the series a_{n_r+1} reduces to zero. For this it is necessary, according to (38.13), that

$$n_r + l + 1 - \frac{Z}{\beta} = 0. \quad (38.15)$$

Since n_r is an integer (including zero), the sum $(n_r + l + 1)$ is also an integer. We denote it by n ; $n = n_r + l + 1$. The integer n is called the principal quantum number, and n_r is called the radial quantum number. For a fixed value of the angular momentum quantum number l we have

$$n \geq l + 1.$$

It is obvious that relation (38.15) determines the ordering of the energy levels of the system. Taking into account the value of β , we find

$$\epsilon = \frac{Z^2}{2n^2}. \quad (38.16)$$

* V.I. Smirnov, *A course of higher mathematics* (Pergamon Press, Oxford, 1964).

Passing over from atomic units to ordinary units, (38.4) and (38.5), we obtain

$$E_n = -\frac{\mu e^4 Z^2}{2\hbar^2 n^2} = -13.5 \frac{Z^2}{n^2} \text{ eV} . \quad (38.17)$$

This formula, first obtained by N. Bohr before the appearance of modern quantum mechanics, determines the discrete energy levels in the hydrogen atom and hydrogen-like ions. We see that the energy levels depend only on the principal quantum number n . The lowest energy level (the ground state) of the particle in the Coulomb field corresponds to the value $n = 1$. The spacing between levels decreases with increasing n , the levels coming nearer to each other. As $n \rightarrow \infty$, $\Delta E \rightarrow 0$ and the discrete spectrum goes over into a continuous one.

The radial function R_{nl} is given by the formula

$$R_{nl} = \text{const } \xi^l e^{-\xi/2} v(\xi) , \quad (38.18)$$

where the polynomial $v(\xi)$, with coefficients determined by the recurrence formula (38.13), coincides except for a constant factor with the generalized Laguerre polynomial*. Hence in our case the radial function assumes the form

$$R_{nl} = A_{nl} \xi^l e^{-\xi/2} L_{n+l}^{2l+1}(\xi) . \quad (38.19)$$

The generalized Laguerre polynomial $L_n^m(\xi)$ is expressed in terms of the derivatives of the Laguerre polynomials which are determined by the relation

$$L_n(\xi) = e^\xi \frac{d^n}{d\xi^n} (e^{-\xi} \xi^n) , \quad (38.20)$$

so that

$$L_n^m(\xi) = \frac{d^m}{d\xi^m} L_n(\xi) . \quad (38.21)$$

The coefficients A_{nl} of (38.19) are determined from the normalization condition**. The radial wave functions belonging, for example, to the two lowest energy levels have the form

$$R_{10}(\rho) = 2 \left(\frac{Z^3}{a^3} \right)^{\frac{1}{2}} e^{-Z\rho} , \quad (38.22)$$

* See the reference on p. 138.

** The calculation of the normalization integral is carried out, for example, in the book of L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

$$R_{20}(\rho) = \left(\frac{Z^3}{2a^3} \right)^{\frac{1}{2}} e^{-\frac{1}{2}Z\rho} (1 - \frac{1}{2}Z\rho), \quad (38.23)$$

$$R_{21}(\rho) = \left(\frac{Z^3}{6a^3} \right)^{\frac{1}{2}} e^{-\frac{1}{2}Z\rho} \frac{1}{2}Z\rho. \quad (38.24)$$

Here the variable ρ (see (38.9)) is again introduced instead of ξ . We stress that the wave function is determined by the whole set of values of the three quantum numbers n , l , and m , whereas the energy levels (38.17) depend only on the principal quantum number n . Thus the energy levels of the hydrogen atom are degenerate. We have seen in §35 that degeneracy in the magnetic quantum number m is a general property of motion in a centrally symmetric field. However, in a Coulomb field the energy levels turn out to also be degenerate in the angular momentum quantum number l . This degeneracy is characteristic only for motion in a Coulomb field. A slight change in the law of force and the energy becomes dependent on the angular momentum quantum number. Hence the degeneracy characteristic of the Coulomb field is called an accidental degeneracy. Let us find the multiplicity of the degeneracy of the n th energy level. Since for a given n the angular momentum quantum number runs over all integers from 0 up to $n-1$ and, in its turn, to each l there correspond $2l+1$ possible values of the quantum number m , the degeneracy is equal to

$$\sum_{l=0}^{n-1} (2l+1) = n^2. \quad (38.25)$$

To each energy level E_n there belong n^2 different wave functions.

Let us consider in more detail the energy levels of the hydrogen atom. They are given by formula (38.17), in which one must set $Z = 1$. The energy of the ground state determines the ionization potential of the hydrogen atom. According to the quantum theory of light emission, the differences between the energy states determine the frequency of electromagnetic waves emitted by the atom (see §103):

$$\hbar\omega = E_m - E_n. \quad (38.26)$$

The quantity E_n/\hbar is called the spectral term. The differences between these spectral terms determine the frequencies of radiation. Substituting expression (38.17) into formula (38.26), we obtain

$$\nu = \frac{\omega}{2\pi} = R \left(\frac{1}{n^2} - \frac{1}{m^2} \right), \quad m > n. \quad (38.27)$$

The quantity R is called the Rydberg constant

$$R = \frac{e^4 \mu}{4\pi\hbar^3} = 3.27 \times 10^{15} \text{ sec}^{-1}. \quad (38.38)$$

All frequencies referring to transitions to one and the same lower level form a spectral series. Thus if we set $n = 1$ in formula (38.27), we obtain the Lyman series. It lies in the ultraviolet part of the spectrum. The transitions to the level $n = 2$ lie in the visible part of the spectrum. The whole set of these spectral lines forms the Balmer series. The spectral series corresponding to transitions to the levels $n = 3$ and so on lie in the infrared region of the spectrum. For hydrogen-like ions the corresponding spectral lines are shifted toward shorter wavelengths, because the frequencies increase by a factor of Z^2 .

Further, we find the probability (35.10) of observing the electron in different quantum states at a given distance r from the nucleus. The ground state of the electron in the hydrogen atom is described by the wave function $\psi_{100} = R_{10}Y_{00}$. For $l = 0, m = 0$ the angular part of the wave function reduces to a constant (see §30), i.e. the state is spherically symmetric. The probability of observing the electron in the ground state ψ_{100} at a given distance from the nucleus is given by the expression

$$dW_{10} = |\psi_{100}|^2 4\pi r^2 dr.$$

Making use of (38.22), we obtain

$$dW_{10} = \frac{4}{a^3} e^{-2r/a} r^2 dr. \quad (38.29)$$

We see that the probability is different from zero over all space, although it decreases rapidly with increasing r . A simple calculation shows that the curve dW_{10}/dr has a maximum at the distance $r = a$, where the quantity a is determined by formula (38.3) and is called the Bohr radius. The form of the function $|R_{nl}|^2 r^2$ for different n and l is shown in fig. V.12. The distance from the centre $\rho = r/a$ is measured along the abscissa, and the probability density $a^3 |R_{nl}|^2 \rho^2$ is measured along the ordinate. We note that the number of zeros of the radial wave function R_{nl} is equal to the value of the radial quantum number n_r . At large distances the radial wave function has the form

$$R_{nl}(r) \sim e^{-Zr/na} \left(\frac{2Zr}{na} \right)^{n-1} + \dots \quad (38.30)$$

The probability density calculated by means of this function rapidly decreases at distances above the order of magnitude of na/Z . Hence it is seen that the

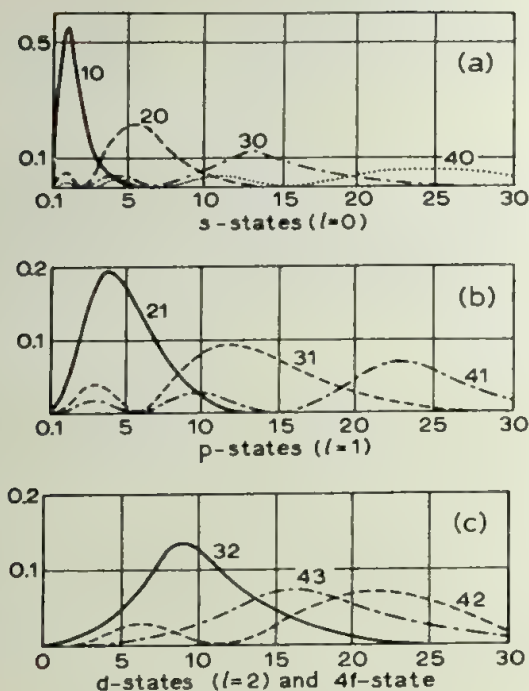


Fig. V.12

quantity na/Z characterizes the size of the atom, because the probability of observing the electron at larger distances is very small.

Up to now we have considered the bound states of an electron in the Coulomb field of the nucleus. Other negatively charged particles, for example π -mesons and muons, can also be in a bound state in the Coulomb field. As we have already mentioned, such systems are called mesic atoms.

As the simplest example let us consider the mesic atom of hydrogen or, as it is called, mesohydrogen. The energy levels of mesohydrogen and the wave functions of the meson are given by formulae (38.17)–(38.19) in which, however, the reduced mass μ of the electron must be replaced by the reduced mass μ' of the meson. The effective size of the atom of mesohydrogen is determined by the value of $a' = (\mu/\mu')a$, which is substantially smaller than the effective size of the hydrogen atom. In particular, the mass of the π^- -meson is equal to 273 electron masses and, correspondingly, $a' \approx 0.2 \times 10^{-10}$ cm. In the mesohydrogen atom the π -meson is situated at considerably

smaller distances from the nucleus than the electron. The presence of the nuclear interaction of the π -meson with the nucleus leads to a displacement of the energy levels (38.17) obtained for the pure Coulomb field. Experimental investigation of this displacement allows one to draw certain conclusions on the character of the nuclear interaction of π -mesons and nucleons. It should be noted that the lifetime of mesic atoms is restricted by the lifetime of the mesons themselves. As is known, mesons are unstable particles undergoing decay with a mean lifetime τ which is characteristic of the given kind of meson.

Up to now we have restricted ourselves to the consideration of the discrete energy spectrum, i.e. we have considered the energy to be negative.

Let us now consider the continuous energy spectrum $E > 0$, $\epsilon = -E/E_0 < 0$ (38.5). We introduce the following notation taking into account (38.7), (38.9) and (38.15):

$$\beta = (2\epsilon)^{\frac{1}{2}} = i(2E/E_0)^{\frac{1}{2}} = ik, \quad n = Z/\beta = -iZ/k, \quad \xi = 2ik\rho. \quad (38.31)$$

Making use of (38.7), (38.14) and (38.31), we write the radial wave function of the continuous spectrum in the form

$$R_{kl} = \frac{C_k}{(2l+1)!} (2k\rho)^l e^{-ik\rho} F\left(1+l+i\frac{Z}{k}; 2l+2, 2ik\rho\right), \quad (38.32)$$

Here C_k is a normalization factor.

If the functions R_{kl} are normalized to the δ -function in k , then this factor is equal to

$$C_k = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{k}{Z} e^{\pi Z/2k} \left| \Gamma\left(l+1 - \frac{iZ}{k}\right) \right|. \quad (38.33)$$

The asymptotic expression of the radial wave function for large ρ is determined by the formula*

$$R_{kl} \approx \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{1}{Z\rho} \sin\left(k\rho + \frac{Z}{k} \ln 2k\rho - \frac{\pi}{2}l + \delta_l\right), \quad (38.34)$$

where

$$\delta_l = \arg \Gamma\left(l+1 - i\frac{Z}{k}\right)$$

* See L.D. Landau and E.M. Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965); V.A. Fok, *Nachala kvantovoi mekhaniki* (Principles of quantum mechanics), (KUBUCH, 1932) p. 155.

(Γ is the gamma-function of a complex variable. Its argument is equal to δ_l).

The expression of the wave function (38.34) (Coulomb field) differs from the general asymptotic expression of the radial wave function in a centrally symmetric field (35.26) by the presence of the slowly increasing logarithmic term in the argument of the sine.

The Quasi-classical Approximation

§39. The limiting transition to classical mechanics

We have more than once referred to the existence of the correspondence principle and the rules for the transition of the relations of quantum mechanics into the formulae of classical mechanics for $\hbar \rightarrow 0$. We shall now define more precisely the conditions of this transition and we shall at the same time obtain an important approximate method of solving the Schrödinger equation* (the WKB method).

If one sets $\hbar = 0$ in the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \left(-\frac{\hbar^2}{2m} \nabla^2 + U \right) \psi, \quad (39.1)$$

it becomes meaningless. Hence to carry out the limiting transition mentioned above we write the wave function ψ in the form

$$\psi = e^{(i/\hbar)S}. \quad (39.2)$$

* The Wentzel-Kramers-Brillouin method. G.Wentzel, Z. Phys. 38 (1926) 518; L.Brillouin, Comptes Rendus 183 (1926) 24, J. de Physique 7 (1926) 353; H.A.Kramers, Z. Phys. 39 (1926) 828, J.Jeffreys, Proc. London Math. Soc. (2) 23 (1923) 428.

Substituting this expression into eq. (39.1), we obtain an equation for the function S :

$$-\frac{\partial S}{\partial t} = \frac{1}{2m} (\nabla S)^2 - \frac{i\hbar}{2m} \nabla^2 S + U, \quad (39.3)$$

We now formally expand the function S in powers of \hbar/i

$$S = S_0 + \left(\frac{\hbar}{i}\right) S_1 + \left(\frac{\hbar}{i}\right)^2 S_2 + \dots \quad (39.4)$$

We substitute the expansion (39.4) into eq. (39.3) and equate the coefficients of the same powers of \hbar . We obtain two equations

$$-\frac{\partial S_0}{\partial t} = \frac{1}{2m} (\nabla S_0)^2 + U, \quad (39.6)$$

$$-\frac{\partial S_1}{\partial t} = \frac{1}{m} \nabla S_0 \nabla S_1 + \frac{1}{2m} \nabla^2 S_0. \quad (39.6)$$

to within terms proportional to the first power of \hbar . Equation (39.5) is the same as the Hamilton–Jacobi equation of classical mechanics* for the action function S_0 . This means that in the zeroth order approximation the motion of the particle follows the classical trajectory. To elucidate the meaning of eq. (39.6) we write the expression for the probability density of finding the particle at a given point of space in the form

$$\rho = |\psi|^2 = e^{2S_1}. \quad (39.7)$$

Multiplying (39.6) by ρ and taking into account that

$$\frac{\partial \rho}{\partial t} = 2 \frac{\partial S_1}{\partial t} \rho; \quad \nabla \rho = 2 \nabla S_1 \rho,$$

we obtain

$$-\frac{\partial \rho}{\partial t} = \frac{1}{m} (\nabla S_0 \nabla \rho + \rho \nabla^2 S_0) = \nabla \cdot \left(\frac{1}{m} \rho \nabla S_0 \right). \quad (39.8)$$

Equation (39.8), equivalent to eq. (39.6), represents a continuity equation. It shows that the probability density moves in space with the same velocity $\mathbf{v} = m^{-1} \nabla S_0$ and on the same trajectory as the particle would move in classi-

* For the Hamilton–Jacobi equation in classical mechanics see L.D.Landau and E.M.Lifshitz, *Mechanics* (Pergamon Press, Oxford, 1960); H.Goldstein, *Classical mechanics* (Addison-Wesley, Cambridge, Mass., 1950).

cal mechanics. We note that, since the velocity is directed along the normal to the surfaces $S_0 = \text{const}$, the trajectories of the classical particle are orthogonal to the surfaces $S_0 = \text{const}$. In the quasi-classical approximation it is natural to call the surfaces $S = \text{const}$ the equi-phase surfaces of the wave function.

We now find the wave function of the stationary states of the particle in the quasi-classical approximation, confining ourselves to one-dimensional motion, so that $\psi = \psi(x, t)$. Because of the stationary state we have

$$\psi(x, t) = e^{-(i/\hbar)Et} \psi(x). \quad (39.9)$$

In correspondence with this, in formulae (39.2) and (39.4) we set

$$S_0(x, t) = -Et + S'_0(x), \quad (39.10)$$

while the functions S_1, S_2, \dots and so on can be assumed to be independent of time.

From eq. (39.5) we obtain

$$E = \frac{1}{2m} \left(\frac{dS'_0}{dx} \right)^2 + U(x), \quad (39.11)$$

whence

$$S'_0(x) = \pm \int [2m(E - U(x))]^{\frac{1}{2}} dx = \pm \int p(x) dx, \quad (39.12)$$

where

$$p(x) = [2m(E - U(x))]^{\frac{1}{2}}.$$

As was to be expected, we have obtained the ordinary formulae of classical mechanics.

We can now determine the function S_1 from eq. (39.6). Taking into account that it is constant in time, we find

$$\frac{dS'_0}{dx} \frac{dS_1}{dx} + \frac{1}{2} \frac{d^2 S'_0}{dx^2} = 0 \quad (39.13)$$

or

$$\frac{dS_1}{dx} = -\frac{1}{2} \frac{d^2 S'_0/dx^2}{dS'_0/dx} = -\frac{1}{2p} \frac{dp}{dx}. \quad (39.14)$$

On integrating we obtain

$$S_1 = -\frac{1}{2} \ln p \quad (39.15)$$

(we shall take the integration constant into account directly in the expression for the wave function).

From the definition (39.2) and the expressions (39.10) and (39.15) we easily find the wave function of the particle with an accuracy to within terms up to the first order in powers of \hbar/i , for $E > U$ and $E < U$ respectively:

$$\psi(x) = \frac{C_1}{[p(x)]^{\frac{1}{2}}} \exp\left(\frac{i}{\hbar} \int p(x) dx\right) + \frac{C_2}{[p(x)]^{\frac{1}{2}}} \exp\left(-\frac{i}{\hbar} \int p(x) dx\right), \quad (39.16)$$

$$\psi(x) = \frac{C_1'}{[|p(x)|]^{\frac{1}{2}}} \exp\left(\frac{1}{\hbar} \int |p(x)| dx\right) + \frac{C_2'}{[|p(x)|]^{\frac{1}{2}}} \exp\left(-\frac{1}{\hbar} \int |p(x)| dx\right), \quad (39.16')$$

The character of the wave function obtained depends critically on the sign of the difference $(E - U)$. If $E > U$, then the momentum is real. This corresponds to a motion of the particle in the region allowed by classical mechanics. In this case the wave function has the character of an oscillatory function. The period of oscillation is smaller, the larger the value of the momentum p . The factor $p^{-\frac{1}{2}} \sim v^{-\frac{1}{2}}$ has a simple meaning. The probability of finding the particle in a region from x to $x+dx$ is proportional to the time during which the particle is in this region; $|\psi(x)|^2 dx \sim v^{-1} dx \sim dt$, i.e. the same result is obtained as in classical mechanics. The wave function in the region of forbidden energies, for $E < U$, has a completely different character. Here the momentum becomes imaginary, and the wave function goes over into a sum of exponential expressions. At the point $E = U$ (called the turning point) $p = 0$ and the expression obtained for the wave function is meaningless.

As is clear from what follows, the quasi-classical approximation becomes inapplicable near the turning point. However, without knowing the wave function at the turning point one cannot close the wave function at the boundary of the allowed and forbidden regions. In other words, one cannot determine the constants figuring in the oscillating and exponential expressions, and without this the quasi-classical wave functions have no practical validity. However, before considering the calculation of the wave function at the turning point it should be explained why the quasi-classical solution is meaningless at this point. For this we estimate the limits of applicability of the expressions (39.16) and (39.16'). First of all we note that in substituting $S = S_0$ into eq. (39.3) we have dropped the term $(\hbar/2m) \nabla^2 S_0$ as negligible. For this to be correct the following inequality must be satisfied

$$\left| \frac{i\hbar}{2m} \nabla^2 S_0 \right| \ll \frac{1}{2m} (\nabla S_0)^2, \quad (39.17)$$

or, taking into account that $\nabla S_0 = \mathbf{p}$,

$$\hbar |\nabla \mathbf{p}| \ll \mathbf{p}^2. \quad (39.18)$$

For one measurement the above inequality can be rewritten in the form

$$\hbar \left| \frac{dp}{dx} \right| \ll p^2. \quad (39.19)$$

Introducing the wave number k instead of the momentum p and the corresponding wavelength $\lambda = \hbar/p = k^{-1}$, we have

$$\frac{d\lambda}{dx} \ll 1. \quad (39.20)$$

Thus we see that the Schrödinger equation reduces to eqs. (39.5) and (39.6) when the condition (39.20) is satisfied. Namely, for the applicability of the quasi-classical approximation it is necessary that the de Broglie wavelength should change sufficiently slowly from point to point in space. In other words, the relative change of the wave number over the extent of a wavelength must be small in comparison with unity $\lambda k^{-1} |dk/dx| \ll 1$. We note also that the relative change in the derivative of the wave number k over the extent of a de Broglie wavelength must be small. Indeed, in obtaining eqs. (39.5) and (39.6) we have also made use of the condition

$$\hbar |\nabla^2 S_1| \ll |\nabla^2 S_0|. \quad (39.21)$$

Taking into account that the problem is one-dimensional and introducing the wave number k , we rewrite this inequality in the form

$$\left| \frac{d^2 k}{dx^2} \right| \lambda \ll \left| \frac{dk}{dx} \right|. \quad (39.22)$$

For the motion of a particle in a potential field $U(x)$ it is convenient to express the wave number in terms of the potential energy according to the formula

$$k = \frac{p}{\hbar} = \left(\frac{2m}{\hbar^2} (E - U) \right)^{\frac{1}{2}}$$

and to write the condition (39.20) in the form

$$\left| \frac{m\hbar}{p^3} \frac{dU}{dx} \right| \ll 1 \quad (39.23)$$

or

$$\frac{m}{\hbar^2} \left| \lambda^3 \frac{dU}{dx} \right| \ll 1. \quad (39.24)$$

Making use of (39.22) one can also write the corresponding inequality for the second derivative of the function U . Hence it is seen that the quasi-classical approximation is valid:

- (1) when the de Broglie wavelength is sufficiently small (i.e. when the particle is moving sufficiently quickly;
- (2) for a sufficiently slow change in the potential energy from point to point, when no considerable change in the momentum of the particle takes place over a length of the order of magnitude of λ .

It becomes clear from (39.23) why the quasi-classical expression for the wave function makes no sense at the turning point. Near the turning point the momentum of the particle becomes small and the quasi-classical approximation becomes inapplicable.

The formulations 'sufficiently small' and 'sufficiently slow' in (1) and (2) stress the fact that, since the criteria (39.20)–(39.24) involve the mass of the particle and the actual characteristics of the field – the quantity dU/dx – then for different fields and different particles the quasi-classical approximation will be valid for motions with different energies. For qualitative estimates one can rewrite (39.20) in a simplified form. Namely, assuming that the change in the wavelength takes place in a region of action of the field having extent a , one can write instead of (39.20) $\lambda \ll a$ or

$$E \gg \frac{\hbar^2}{2ma^2}. \quad (39.25)$$

For α -particles ($m = 6.7 \times 10^{-24}$ g) with energy $E = 1$ MeV passing through an atomic shell ($a \sim 10^{-8}$ cm) the inequality (39.25) is fulfilled to a good approximation. On the contrary, for the same α -particles with an energy of 10 MeV undergoing direct collision with a nucleus ($a \sim 10^{-13}$ cm) the quasi-classical consideration is inapplicable. In the region of substantially larger energies the application of the quasi-classical approximation turns out to be possible in considering certain processes connected with nuclear collisions.

§40. The solution of the Schrödinger equation near a turning point

Let us now return to the consideration of the behaviour of the wave function near a turning point.

The idea here is as follows: since near a turning point the quasi-classical approximation turns out to be unsatisfactory, it is necessary to find a solution of the Schrödinger equation without making use of this approximation. The possibility of obtaining such a solution for an arbitrary form of the potential energy is associated with the fact that the expression for the potential energy near a turning point makes possible an essential simplification (see below). If the solution sought is found, then one has to determine its asymptotic behaviour at large distances from the turning point in both directions, in those regions where the quasi-classical approximation is already valid. Requiring that the quasi-classical solution be the same as this asymptotic expression, we shall be able to determine the corresponding constants.

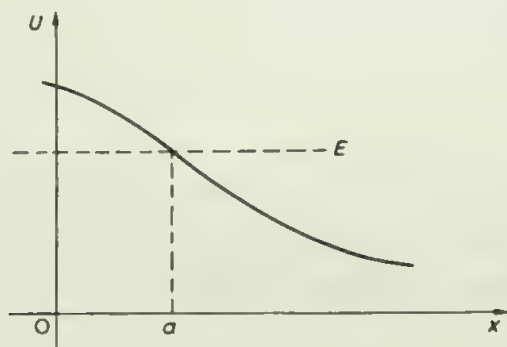


Fig. V.13

To carry out this programme, we note that one can expand the potential energy $U(x)$ in the vicinity of the turning point (fig. V.13) in a series with respect to the small displacement $\xi = x - a$ and retain the linear term of this expansion. In this case we assume that at the turning point the curve $U(x)$ is smooth, as is shown in fig. V.13. We shall also assume that the region $x > a$ extends to infinity. Near the point $x = a$ we can write

$$U(x) = U(a) + \left. \frac{dU}{dx} \right|_{x=a} (x - a) + \dots \quad (40.1)$$

The potential energy at the point a is the same as the total energy of the particle $U(a) = E$. We denote by f the force acting on the particle at the turning point, $f = -dU/dx|_{x=a}$, and introduce the new variable $\xi = x - a = (E - U)/f$. We write the Schrödinger equation near the point $x = a$ as

follows:

$$\frac{d^2\psi}{d\xi^2} + f \frac{2m}{\hbar^2} \xi \psi = 0. \quad (40.2)$$

The Schrödinger equation was considered in such a form in § 13. Equation (40.2) is the same as eq. (13.5) for $E = 0$. Consequently, the wave function satisfying eq. (40.2) and finite for $\xi \rightarrow \pm \infty$ is expressed in terms of the Airy function. We shall use immediately the asymptotic expressions (13.9) and (13.10). This means that we consider values of ξ which are sufficiently large for the asymptotic expressions to be used, and which are at the same time such that expansion (40.1) is still applicable. Such a region, as a rule, exists for fields satisfying the quasi-classical conditions.

Correspondingly, in the region $x \gg a$ the solution of eq. (40.2) can be written in the form

$$\psi = \frac{2C}{(2mf\xi)^{\frac{1}{4}}} \sin \left(\frac{2}{3\hbar} (2mf)^{\frac{1}{2}} \xi^{\frac{3}{2}} + \frac{1}{4}\pi \right) \quad (40.3)$$

where C is the normalization constant.

For motion in the field (40.1) the momentum p is of the form

$$p = [2m(E - U)]^{\frac{1}{2}} = (2mf\xi)^{\frac{1}{2}}. \quad (40.4)$$

We express the action in terms of the variable ξ

$$\int_a^x p dx = \int_0^\xi p d\xi = (2mf)^{\frac{1}{2}} \int_0^\xi \xi^{\frac{1}{2}} d\xi = \frac{2}{3} (2mf\xi^3)^{\frac{1}{2}}. \quad (40.5)$$

Making use of (40.5) we can write the wave function (40.3) in the form

$$\psi(x) = 2Cp^{-\frac{1}{2}} \sin \left(\hbar^{-1} \int_a^x p dx + \frac{1}{4}\pi \right) = 2Cp^{-\frac{1}{2}} \cos \left(\hbar^{-1} \int_a^x p dx - \frac{1}{4}\pi \right). \quad (40.6)$$

We see that the function (40.6) has the quasi-classical form (see (39.16)).

We now find the function in the region $x \ll a$. Again making use of the asymptotic expression (13.9) and expressions (40.4), and (40.5), we have

$$\psi = \frac{C}{(2mf\xi)^{\frac{1}{4}}} \exp \left[-\frac{2}{3}\hbar^{-1} (2mf\xi^3)^{\frac{1}{2}} \right] = \frac{C}{|p|^{\frac{1}{2}}} \exp \left[-\hbar^{-1} \int_x^a |p| dx \right], \quad (40.7)$$

where C is the same normalization constant as in formula (40.6). Thus we

have obtained the expression for the quasi-classical wave function valid on both the left and right side of the turning point $x = a$.

We finally write the expressions for the quasi-classical wave function:

$$\psi(x) = \begin{cases} C|p|^{-\frac{1}{2}} \exp \left(-\hbar^{-1} \int_x^a |p| dx \right) & (x < a), \\ 2Cp^{-\frac{1}{2}} \cos \left(\hbar^{-1} \int_a^x p dx - \frac{1}{4}\pi \right) & (x > a). \end{cases} \quad (40.8)$$

The constant C is defined by the normalization condition.

Analogously, if the allowed region lies on the left of the turning point b , i.e. $U(x) < E$ for $x < b$ and $U(x) > E$ for $x > b$ (see, for example, fig. V.14 for $x > a$), then the wave function is written in the form

$$\psi(x) = \begin{cases} C'|p|^{-\frac{1}{2}} \exp \left(-\hbar^{-1} \int_b^x |p| dx \right) & (x > b), \\ 2C'p^{-\frac{1}{2}} \cos \left(\hbar^{-1} \int_x^b p dx - \frac{1}{4}\pi \right) & (x < b). \end{cases} \quad (40.9)$$

Thus we have found in the quasi-classical approximation the function $\psi(x)$ satisfying the Schrödinger equation. The solution obtained is still not complete, since a linear equation of second order has two linearly independent solutions. In the case of a wave function depending on one independent variable the other solution of the Schrödinger equation can easily be obtained. That is, if ψ_1 and ψ_2 are two linearly independent functions satisfying the one-dimensional Schrödinger equation corresponding to an energy E , then they are always connected by the relation

$$\frac{1}{\psi_1} \frac{d^2 \psi_1}{dx^2} = \frac{1}{\psi_2} \frac{d^2 \psi_2}{dx^2} = \frac{2m}{\hbar^2} (U - E). \quad (40.10)$$

Integrating, we obtain

$$\psi_2 \frac{d\psi_1}{dx} - \psi_1 \frac{d\psi_2}{dx} = \text{const}. \quad (40.11)$$

We seek the solution linearly independent of (40.8) in the form

$$\psi(x) = \begin{cases} B_1 |p|^{-\frac{1}{2}} \exp \left(\hbar^{-1} \int_x^a |p| dx \right) & (x < a), \\ B_2 p^{-\frac{1}{2}} \cos \left(\hbar^{-1} \int_a^x p dx + \alpha \right) & (x > a). \end{cases} \quad (40.12)$$

The expressions (40.8) and (40.12) are to be substituted into relation (40.11). Then, by virtue of inequality (39.20), it is sufficient to restrict oneself to differentiation with respect to the arguments of the exponential and trigonometric functions. Equating the expressions obtained for $x < a$ and $x > a$, we find $B_1 = B_2 \sin(\frac{1}{4}\pi + \alpha)$. Finally we set $B_1 = B_2$, $\alpha = \frac{1}{4}\pi$. Thus the solution linearly independent of (40.8) can be chosen in the form

$$\psi(x) = \begin{cases} B |p|^{-\frac{1}{2}} \exp \left(\hbar^{-1} \int_x^a |p| dx \right) & (x < a), \\ B p^{-\frac{1}{2}} \cos \left(\hbar^{-1} \int_a^x p dx + \frac{1}{4}\pi \right) & (x > a). \end{cases} \quad (40.13)$$

Correspondingly, the solution linearly independent of (40.9) is written as

$$\psi(x) = \begin{cases} B' |p|^{-\frac{1}{2}} \exp \left(\hbar^{-1} \int_b^x |p| dx \right) & (x > b), \\ B' p^{-\frac{1}{2}} \cos \left(\hbar^{-1} \int_x^b p dx + \frac{1}{4}\pi \right) & (x < b). \end{cases} \quad (40.14)$$

The expressions obtained will be unsuitable in the case where at the turning point, for example at point b , the potential energy becomes infinite in a discontinuous way. In this case $\psi = 0$ in the region $x \geq b$. The phase of the wave function for $x < b$ can be determined if the conditions of applicability of the quasi-classical approximation (39.20) remain valid up to the point $x = b$. Then, taking into account that $\psi(b) = 0$, we obtain

$$\psi(x) = A p^{-\frac{1}{2}} \sin \left(\hbar^{-1} \int_x^b p dx \right). \quad (40.15)$$

§41. Motion in a potential well in the quasi-classical approximation

We apply the results obtained to the motion of a particle in a potential well. We then find an approximate formula for the energy spectrum. Its comparison with accurate formulae will clearly allow us to judge the degree of accuracy and the merits of the quasi-classical approximation. At the same time the solution of the problem posed is of great interest in another respect. It makes it possible to elucidate the connection between quantum mechanics and the old Bohr theory.

Let us consider, first of all, a potential well with infinitely high walls (see §8). The wave function in the quasi-classical approximation is given by a formula of the type (40.15). Namely,

$$\psi(x) = Ap^{-\frac{1}{2}} \sin \left(\hbar^{-1} \int_a^x p dx \right). \quad (41.1)$$

In the potential well there will be two turning points a and b at which the wave function must reduce to zero. Thus at the two turning points the condition $\psi = 0$ or

$$Ap^{-\frac{1}{2}} \sin \left(\hbar^{-1} \int_a^b p dx \right) = 0 \quad (41.2)$$

must be fulfilled. This condition is fulfilled if

$$\hbar^{-1} \int_a^b p dx = n\pi, \quad (41.3)$$

where the n are the positive integers beginning with unity. Since the momentum is constant and equal to $(2mE)^{\frac{1}{2}}$, we find

$$E = \frac{\hbar^2 \pi^2}{2ml^2} n^2, \quad (41.4)$$

where $l = b - a$ is the width of the well.

We see that in the simplest case of an infinitely deep potential well the quasi-classical approximation leads to an accurate expression for the energy spectrum (see §8).

Let us now consider the general case of a potential well as shown in fig. V.14. We assume that the forbidden region extends infinitely to the right and left of the turning points. Then the quasi-classical wave function will not

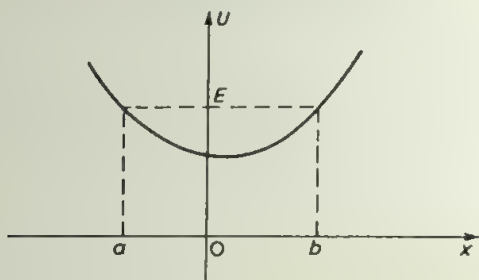


Fig. V.14

contain any exponentially increasing terms and will be given by formulae of the type (40.8) or (40.9). The two wave functions (40.8) and (40.9) describing the motion of a particle in a well must be identical.

$$2Cp^{-\frac{1}{2}} \cos \left(\hbar^{-1} \int_a^x p dx - \frac{1}{4} \pi \right) = 2C'p^{-\frac{1}{2}} \cos \left(\hbar^{-1} \int_x^b p dx - \frac{1}{4} \pi \right) \quad (41.5)$$

$(a < x < b).$

This is possible only in the case where the sum of the two phases is equal to an integer multiple of π

$$\hbar^{-1} \int_a^b p dx - \frac{1}{2} \pi = n\pi, \quad (41.6)$$

where n is an integer. Then

$$C' = (-1)^n C. \quad (41.7)$$

If one introduces the integral with respect to the period of the classical motion of a particle from a to b and conversely $\int_a^b p dx = \frac{1}{2} \oint p dx$, then from (41.6) we obtain

$$\oint p dx = 2\pi\hbar(n + \frac{1}{2}). \quad (41.8)$$

The above expression is none other than the Bohr quantization rule, from which the stationary states of a particle in the quasi-classical case are determined. Thus Bohr's theory with its inconsistent imposition of quantization conditions upon purely classical quantities, turns out to be completely valid within the limits of the quasi-classical approximation. We note that the

number n is equal to the number of roots of the quasi-classical wave function between the turning points a and b , because as x changes from a to b the phase of the wave function increases from $-\frac{1}{4}\pi$ to $(n + \frac{1}{2})\pi - \frac{1}{4}\pi$ and, consequently, the cosine reduces to zero n times. The larger the quantum number n , i.e. the smaller the de Broglie wavelength, the better the conditions of applicability of the quasi-classical approximation (39.20). Consequently, we expect that the energy levels obtained from condition (41.8) coincide, for large values of n , with their exact values as calculated from the solution of the Schrödinger equation. However, in some cases, such as, for example, the harmonic oscillator, formula (41.8) gives the correct value of the energy level for any value of n . The integral on the left-hand side of eq. (41.8) represents the area bounded in its phase plane by the classical phase trajectory of a particle with energy E . According to (41.8), this area is equal to $2\pi\hbar n$ for $n \gg 1$. Since an energy level of the system corresponds to each node of the wave function, the number n gives the number of states with energies less than or equal to E . Thus to each quantum state in the phase plane there corresponds an area equal to $2\pi\hbar$. The number of states corresponding to an area $\Delta p \Delta x$ in the phase plane will thus be equal to

$$\frac{\Delta p \Delta x}{2\pi\hbar}. \quad (41.9)$$

Generalizing this formula to the three-dimensional case it is obvious that we shall obtain the number of states corresponding to the volume $\Delta x \Delta y \Delta z \Delta p_x \Delta p_y \Delta p_z$ in phase space to be

$$\frac{\Delta x \Delta y \Delta z \Delta p_x \Delta p_y \Delta p_z}{(2\pi\hbar)^3}. \quad (41.10)$$

This formula was the basis of our exposition of statistical physics. We see that the quasi-classical approximation is like a bridge connecting classical and quantum mechanics. It enables one to understand the meaning of Bohr's theory and the correspondence principle, and it makes it possible to eliminate all apparent contradictions between the different aspects of the behaviour of real particles. By means of the quasi-classical approximation we can find directly the conditions and the degree of accuracy with which one can pass over to the classical description of the motion of particles in many problems. At the same time it gives a relatively simple method of describing quantum systems approximately, such as, in particular, finding energy levels for particles with a high energy.

§42. Potential barrier penetration

In §13 we considered the passage of a microparticle through a rectangular potential barrier. In this section we shall obtain more general formulae for the case of the passage of particles through potential barriers of arbitrary form (fig. V.15). We assume that the energies E of the particles are sufficiently large, and that the curve of the potential energy is sufficiently smooth that, with the exception only of small regions around the turning points a and b , the conditions of applicability of the quasi-classical approximation are everywhere fulfilled.

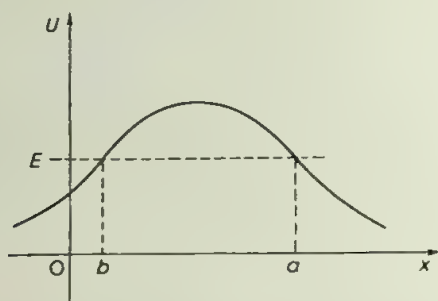


Fig. V.15

Let a particle moving from left to right along the x -axis fall onto a barrier. Then in the region behind the turning point a , i.e. for $x > a$, there must be only a wave propagating in the positive direction of the x -axis (in this region there is no reflected wave). The quasi-classical wave function for $x > a$ can be written in the form of a superposition of the expressions (40.8) and (40.13)

$$\psi(x) = 2Cp^{-\frac{1}{2}} \cos \left(\hbar^{-1} \int_a^x p dx - \frac{1}{4}\pi \right) + Bp^{-\frac{1}{2}} \cos \left(\hbar^{-1} \int_a^x p dx + \frac{1}{4}\pi \right). \quad (42.1)$$

Since at large distances from the point a the momentum p changes little, each term of (42.1) is a superposition of two plane waves propagating in opposite directions. It is easily seen that the superposition (42.1) describes a wave propagating from left to right only under the condition $B = -2Ci$. In this case

$$\psi = 2Cp^{-\frac{1}{2}} \exp \left(i\hbar^{-1} \int_a^x p dx - \frac{1}{4}\pi \right) \quad (x > a). \quad (42.2)$$

Let us find the quasi-classical wave function in the region $b < x < a$. Taking the superposition (40.8) and (40.13) and taking into account the relation between B and C , we obtain

$$\begin{aligned} \psi(x) = C|p|^{-\frac{1}{2}} & \left[\exp \left(-\hbar^{-1} \int_x^a |p| dx \right) + \right. \\ & \left. + 2i^{-1} \exp \left(\hbar^{-1} \int_x^a |p| dx \right) \right] \quad (b < x < a) \quad (42.3) \end{aligned}$$

This relation is conveniently rewritten in the form

$$\begin{aligned} \psi(x) = C|p|^{-\frac{1}{2}} & \left[e^{-L} \exp \left(\hbar^{-1} \int_b^x |p| dx \right) - \right. \\ & \left. - 2ie^L \exp \left(-\hbar^{-1} \int_b^x |p| dx \right) \right] \quad (b < x < a) \quad (42.4) \end{aligned}$$

where

$$L = \hbar^{-1} \int_b^a |p| dx.$$

Making use now of the relations (40.9) and (40.14), one easily obtains the expression for the wave function in the region in front of the barrier

$$\begin{aligned} \psi(x) = Cp^{-\frac{1}{2}} & \left[4i^{-1} e^L \cos \left(\hbar^{-1} \int_x^b p dx - \frac{1}{4} \pi \right) + \right. \\ & \left. + e^{-L} \cos \left(\hbar^{-1} \int_x^b p dx + \frac{1}{4} \pi \right) \right] = \\ = 2i^{-1} Cp^{-\frac{1}{2}} & \left[(e^L - \frac{1}{4} e^{-L}) \exp \left(i\hbar^{-1} \int_x^b p dx - \frac{1}{4} i\pi \right) + \right. \\ & \left. + (e^L + \frac{1}{4} e^{-L}) \exp \left(-i\hbar^{-1} \int_x^b p dx + \frac{1}{4} i\pi \right) \right] \quad (x < b). \quad (42.5) \end{aligned}$$

Thus we have written the wave function in the region in front of the barrier in the form of a superposition of incident and reflected waves. We now determine the transmission coefficient, D , of particles through the barrier. Making use of (42.2), we calculate the current density of particles transmitted through the barrier j_{tr}

$$j_{tr} = 4|C|^2 m^{-1}. \quad (42.6)$$

The current density of incident particles, in correspondence with (42.5), is equal to

$$j_{inc} = 4|C|^2 m^{-1} (e^L + \frac{1}{4}e^{-L})^2. \quad (42.7)$$

Consequently, we obtain the following expression for the transmission coefficient D :

$$D = \frac{j_{tr}}{j_{inc}} = \frac{e^{-2L}}{(1 + \frac{1}{4}e^{-2L})^2}. \quad (42.8)$$

For a sufficiently wide potential barrier $e_b^{-2L} \ll 1$, we obtain

$$D = e^{-2L} = \exp \left(-2\hbar^{-1} \int_a^b p dx \right). \quad (42.9)$$

We note that if the potential energy on one side of the barrier changes rapidly enough, so that the quasi-classical approximation is inapplicable, then in the expression (42.9) a factor will appear before the exponential function. However, the basic exponential factor does not change. Formula (42.9) is widely used for the calculation of the probabilities of transmission of particles through potential barriers.

As an example let us consider the theory of α -decay. It is well known that all heavy nuclei with mass numbers of the order of magnitude of 200 turn out to be unstable with respect to α -decay. The probability of decay strongly depends on the energy of the emitted α -particles and varies over a very wide range. Thus if the decay probability is characterized by a half-life τ , then τ is equal to 1.6×10^{-4} sec in the case of ^{234}Po which emits α -particles with an energy of 7.8 MeV, whereas τ is equal to 1.4×10^{10} years in the case of ^{232}Th which emits α -particles with an energy of 4 MeV. Such a strong dependence of the probability of α -decay on the energy is accounted for by the fact that the particle, in order to get out of the nucleus, must pass through a potential barrier*. Indeed, simplifying the treatment, we can assume that the initial

* R.Gurney and E.Condon, Nature 122 (1928) 439; Phys. Rev. 33 (1929) 127.

nucleus already contains an α -particle. Then the problem reduces to the calculation of the probability that the α -particle will leave the initial, parent nucleus.

We denote by $U(r)$ the energy of interaction of the α -particle with the remaining daughter nucleus. At small distances $U(r)$ amounts to the potential of nuclear forces which we shall consider constant and equal to U_0 , while at large distances it is just the Coulomb interaction between the α -particle and the remaining nucleus (see fig. V.16).

$$U(r) = \begin{cases} 2Ze^2/r & (r > r_0), \\ U_0 & (r \leq r_0), \end{cases}$$

where r_0 is a distance of the order of magnitude of the nuclear size.

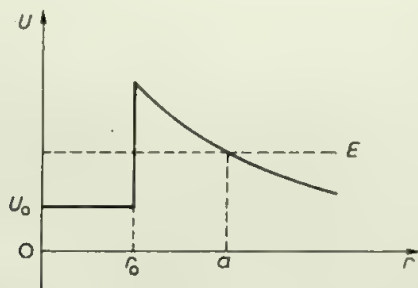


Fig. V.16

Making use of formula (42.9), we can find the probability of the α -particle passing through the potential barrier. The fact that formula (42.9) is derived for one-dimensional motion is of no importance here, because we have established in §35 that a radial motion is equivalent to a one-dimensional motion with a certain effective potential energy.

For simplicity we consider the case where $l = 0$, so that the centrifugal energy is not involved in the calculations (see also the next section). We then have

$$D = \exp \left[-2\hbar^{-1} \int_{r_0}^a \left(2\mu \left(\frac{2Ze^2}{r} - E \right) \right)^{\frac{1}{2}} dr \right] = e^{-2L}, \quad (42.10)$$

where the turning point a is determined by the condition $a = 2Ze^2/E$, and μ is the reduced mass of the α -particle and the daughter nucleus. We calculate the

integral L :

$$L = \left(\frac{4\mu Ze^2}{\hbar^2} \right)^{\frac{1}{2}} \int_{r_0}^a \left(1 - \frac{E}{U_{\max}} \frac{r}{r_0} \right)^{\frac{1}{2}} r^{-\frac{1}{2}} dr,$$

where $U_{\max} = 2Ze^2/r_0$. On substituting $(Er/U_{\max}r_0)^{\frac{1}{2}} = \sin \alpha$, we easily obtain

$$L = \frac{2Ze^2}{\hbar v} (\pi - 2\alpha_0 - \sin 2\alpha_0), \quad (42.11)$$

where $\sin \alpha_0 = (E/U_{\max})^{\frac{1}{2}}$ and $v = (2E/\mu)^{\frac{1}{2}}$ is the velocity of the α -particle, if the difference between the reduced mass μ and the mass of the α -particle is neglected. Thus the probability for the α -particle to pass through the barrier is given by the expression

$$D = \exp \left[-\frac{4Ze^2}{\hbar v} (\pi - 2\alpha_0 - \sin 2\alpha_0) \right]. \quad (42.12)$$

We obtain the probability λ for α -decay if we multiply the probability D of passing through the barrier by the probability for α -decay in the absence of the barrier, $\lambda = \nu D$. The quantity ν cannot be calculated with any accuracy. It is significant, however, that the very strong dependence of the decay probability on the energy of the α -particle is involved in the factor D . Qualitatively this dependence is well confirmed by experiment.

Finally we note that similar reasoning is also applicable in the case of spontaneous fission of heavy nuclei.

§43. Quasi-classical motion in a centrally symmetric field

Let us find an approximate expression for the radial component of the wave function $R(r)$ or the function $\chi(r) = rR$ if the potential energy $U(r)$ satisfies the condition for the quasi-classical approximation. We can make use of the relations already derived, since the function $\chi(r)$, as we know (see §35), is described by the one-dimensional Schrödinger wave equation with the effective potential energy

$$U_{\text{eff}}(r) = U(r) + \frac{\hbar^2 l(l+1)}{2mr^2}.$$

However, in this case, of course, the fact that the coordinate r , as distinct from x , varies from 0 to ∞ must be taken into account. For $l = 0$ we have $U_{\text{eff}}(r) = U(r)$. If the quasi-classical conditions are fulfilled up to the point

$r = 0$, then the corresponding wave function can easily be obtained. Indeed, the condition for the finiteness of the wave function at zero gives $\chi(0) = 0$ and, making use of (40.15), we obtain

$$\chi(r) = C[2m(E - U(r))]^{-\frac{1}{2}} \sin \left(\hbar^{-1} \int_0^r [2m(E - U(r))]^{\frac{1}{2}} dr \right). \quad (43.1)$$

In the more general case $l \neq 0$ the effective potential energy $U_{\text{eff}}(r)$ must satisfy the quasi-classical condition. If it is assumed that at small distances the centrifugal energy $\hbar^2 l(l+1)/2mr^2$ plays the basic role and that $p \sim \hbar l/r$, then it follows from the condition (39.20) that $l \gg 1$. The corresponding wave function for $r \gg a$, where a is the turning point, can be written in the form (40.8) with the condition, however, that in the centrifugal energy the quantity $l(l+1)$ is replaced by* $(l + \frac{1}{2})^2$:

$$\chi(r) = Ap_r^{-\frac{1}{2}} \cos \left(\hbar^{-1} \int_a^r p_r dr - \frac{1}{4}\pi \right) \quad (43.2)$$

where

$$p_r = \left[2m \left(E - U(r) - \frac{\hbar^2 (l + \frac{1}{2})^2}{2mr^2} \right) \right]^{\frac{1}{2}}. \quad (43.3)$$

Thus the term $\hbar^2/8mr^2$ is added to the centrifugal energy. This addition leads to a more correct value of the phase of the wave function. Thus for free motion, $U = 0$, formula (43.2) gives for the phase of the wave function at large distances the value which we have obtained in §35**.

* H.A. Kramers, Z. Phys. 39 (1926) 828.

** See, for example, L.D. Landau and E.M. Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

The Matrix Form of Quantum Mechanics

§44. Operators and matrices

The mathematical apparatus of quantum mechanics developed in the preceding chapters (the method of linear Hermitian operators) is not the only mathematical apparatus used in quantum mechanics. It turns out that all mechanical quantities in quantum mechanics can be related to so-called Hermitian matrices as well as to operators. A matrix R is understood to be the whole set of quantities forming the table

$$R = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1n} \\ R_{21} & R_{22} & \dots & R_{2n} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ R_{n1} & R_{n2} & \dots & R_{nn} \end{bmatrix} \quad (44.1)$$

The number of rows and columns in the table need not, in the general case, be the same. Each of the quantities (generally speaking complex) appearing in the table is called a matrix element. A matrix element has two indices: the

first denotes the ordinal number of the row, and the second denotes the ordinal number of the column. The concept of matrices is usually introduced in connection with the linear transformation of vectors in an n -dimensional space*.

We shall see below that in quantum mechanics it is possible to make a geometrical interpretation of the wave function as a vector in a certain imaginary space. In the meantime we shall convince ourselves by means of very general reasoning of the fact that any linear operator \hat{F} can be related to a matrix F with definite values of the matrix elements.

The definition of an operator means that the result of its action on the function $\psi(x)$ is given by

$$\hat{R}\psi(x) = \varphi(x). \quad (44.2)$$

We pass over from the x -representation to the F -representation. For this we expand the functions $\psi(x)$ and $\varphi(x)$ in terms of the eigenfunctions $\psi_m(x)$ of the operator \hat{F} . We assume that the operator \hat{F} has a discrete spectrum. For example, such a representation can be the energy representation (E -representation)

$$\psi(x) = \sum_m c_m \psi_m(x), \quad \varphi(x) = \sum_n b_n \psi_n(x). \quad (44.3)$$

The whole set of amplitudes c_m (or b_n) determines the wave function ψ (or φ) in the F -representation. Sometimes this set is conveniently denoted in the form of a column

$$c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ \vdots \end{pmatrix}; \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ \vdots \end{pmatrix} \quad (44.4)$$

We substitute the expansion (44.3) into (44.2), and obtain

$$\sum_n b_n \psi_n(x) = \sum_m c_m \hat{R} \psi_m(x).$$

Multiplying the left-hand and right-hand sides of this equality by $\psi_l^*(x)$ and

* For more details see for example V.I. Smirnov, *A course of higher mathematics* (Pergamon Press, Oxford, 1964).

integrating over the entire region of variation of the independent variables, we find

$$b_l = \sum_m R_{lm} c_m, \quad (44.5)$$

where

$$R_{lm} = \int \psi_l^*(x) \hat{R} \psi_m(x) dV. \quad (44.6)$$

The relation (44.5) determines directly the transformation of the function ψ into the function φ in the F -representation under the action of the operator \hat{R} . The operator \hat{R} in this representation is given by formula (44.6), i.e. in the form of a matrix. Thus the definition of the matrix R is equivalent to the definition of the operator \hat{R} itself.

The matrix element R_{ik} is sometimes called the matrix element corresponding to the transition from the k th state into the i th state. Such a terminology is based on the following reasoning. We assume that the initial state of the system is the k th state, $\psi(x) = \psi_k(x)$. Under the action of the operator \hat{R} the transformation (44.2) takes place. Making use of (44.3) and (44.5) and taking into account that in the given case $c_m = \delta_{mk}$, $b_n = R_{nk}$, we obtain

$$\varphi(x) = \hat{R} \psi_k = \sum_n b_n \psi_n = \sum_n R_{nk} \psi_n(x), \quad (44.7)$$

and, consequently, the square of the modulus of the matrix element R_{ik} determines the probability of finding the system in the i th state.

Knowing the matrix corresponding to the quantity R , one can also easily find the mean value of this quantity in a certain state ψ . According to the general formula (22.4) we have

$$\bar{R} = \int \psi^* \hat{R} \psi dV.$$

Substituting here the expansion (44.3) instead of ψ , we obtain

$$\bar{R} = \sum_m \sum_n c_m^* c_n \int \psi_m^* \hat{R} \psi_n dV = \sum_m \sum_n c_m^* R_{mn} c_n. \quad (44.8)$$

We note that if we determine the matrix elements (44.6) by means of the wave functions ψ_m which are the eigenfunctions of the operator \hat{R} , then we also determine the matrix R of the operator in its own representation

$$R_{ml} = \int \psi_m^* \hat{R} \psi_l dV = R_l \int \psi_m^* \psi_l dV = R_l \delta_{ml}. \quad (44.9)$$

We see that in this case only the matrix elements with $m = l$ are different from

zero. Matrices of such a form are called diagonal

$$R = \begin{vmatrix} R_{11} & 0 & 0 & \dots \\ 0 & R_{22} & 0 & \dots \\ 0 & 0 & R_{33} & \dots \\ \dots & \dots & \dots & \dots \end{vmatrix}. \quad (44.10)$$

Thus in its own representation any operator is described by a diagonal matrix, and the diagonal elements are equal to the eigenvalues of this operator.

The definition of the matrix elements R_{ml} is completely equivalent to the definition of the operator \hat{R} . As we shall see, it enables one to determine the eigenvalues and eigenfunctions of this operator. On the other hand, if the operator \hat{R} is known, then also the matrix elements R_{ml} can be determined.

The Hermitian character of the operator \hat{R} imposes a certain restriction upon the form of the matrix elements R_{ml} . Namely, for the matrix element R_{ml}^* , the complex conjugate of the element R_{ml} , we have

$$R_{ml}^* = \left(\int \psi_m^* \hat{R} \psi_l dV \right)^* = \int \psi_m \hat{R}^* \psi_l^* dV.$$

By definition of the Hermitian operator

$$\int \psi_m \hat{R}^* \psi_l^* dV = \int \psi_l^* \hat{R} \psi_m dV,$$

so that

$$R_{ml}^* = R_{lm}.$$

We see that the Hermitian property of matrix elements

$$R_{lm} = R_{ml}^*, \quad (44.11)$$

follows from the requirement of the Hermitian character of the operator.

Every physical quantity in quantum mechanics, as well as the Hermitian operator \hat{R} , can be related to a Hermitian matrix R whose matrix elements are determined by formula (44.6).

As we shall see below, the matrix form of quantum mechanics is, in certain cases, more convenient than the operator form. The representation of quantum mechanics in the matrix form will enable us to formulate the equations of quantum mechanics like the equations of classical physics. The wave function will no longer appear in them. The equations in this form will coincide with those of classical mechanics, with only the basic difference that

in these equations classical quantities will be replaced by the corresponding matrices. However, before passing over to a systematic exposition of quantum mechanics in the matrix form it is necessary to present the basic notions of matrix calculus.

§45. The fundamentals of matrix calculus

In the preceding section we have defined an arbitrary matrix R as the whole set of quantities R_{bm} arranged in a definite order in the form of the table

$$R = \begin{vmatrix} R_{11} & R_{12} & \dots & & \\ R_{21} & R_{22} & \dots & & \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & \cdot & \\ R_{n1} & R_{n2} & \dots & R_{nn} & \dots \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & \cdot & \end{vmatrix} \quad (45.1)$$

Matrices in which the number of columns is equal to the number of rows are called square matrices. A matrix can be finite, if the number of columns and the number of rows is finite, as well as infinite if the number is arbitrarily large. The matrix elements $R_{11}, R_{22}, \dots, R_{nn} \dots$ forming the diagonal of the matrix are called the diagonal elements. The matrix for which all elements are equal to zero is called the null matrix O . The matrix for which all the diagonal elements are equal to unity is called the unit matrix. We shall denote this matrix by

$$I = \begin{vmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{vmatrix}, \quad (45.2)$$

i.e. $(I)_{mn} = \delta_{mn}$. A matrix can be considered as a certain hypercomplex number, just as the set of two numbers a and b can be treated as one complex number $a + ib$. As in the case of complex numbers, one can construct an

algebra of hypercomplex numbers (matrices) by defining the action of addition and multiplication of matrices.

The matrix R , each element of which is equal to the sum of the corresponding matrix elements of the matrices F and D ,

$$\begin{array}{l} R = F + D, \\ \text{if} \\ R_{ml} = F_{ml} + D_{ml}, \end{array} \quad (45.3)$$

is called the sum of the two matrices F and D . One can add or subtract only matrices having the same number of rows and columns. Two matrices F and D are equal to each other if the corresponding matrix elements are equal to each other:

$$\begin{array}{l} F = D, \\ \text{if} \\ F_{ml} = D_{ml}. \end{array} \quad (45.4)$$

Further we define the product of the number k and the matrix D as a matrix F each element of which is equal to the product of the number k and the corresponding matrix element of the matrix D :

$$\begin{array}{l} F = kD, \\ F_{ml} = kD_{ml}. \end{array} \quad (45.5)$$

The matrix L is called the product of the matrices F and D

$$L = FD,$$

if the matrix element L_{mn} is equal to

$$L_{mn} = \sum_l F_{ml} D_{ln}. \quad (45.6)$$

This means that each element of the matrix L is equal to the sum of the products of the elements of the m th row of the matrix F and the elements of the n th column of the matrix D .

The matrix F can be multiplied by the matrix D only in the case where the number of columns of the matrix F is equal to the number of rows of the matrix D . We stress that the product of matrices, as the product of operators, is non-commuting, i.e. in the general case

$$FD \neq DF.$$

If the unit matrix (45.2) is taken as one of the factors, then we arrive at the equalities

$$F1 = 1F = F, \quad (45.7)$$

i.e. multiplication by the unit matrix is commutative. We note that when two matrices with matrix elements different from zero are multiplied a null matrix can be obtained. Let, for example,

$$F = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}; \quad D = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}.$$

Then

$$L = FD = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0.$$

On the other hand

$$L' = DF = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} \neq 0.$$

Analogously to (45.6) one can form the product of three and more matrices. Thus if $FD = L$, then the product RFD is equal to

$$(RFD)_{mn} = \sum_l R_{ml} L_{ln} = \sum_l \sum_p R_{ml} F_{lp} D_{pn}. \quad (45.8)$$

We arrive at the exactly the same expression if we take the product of the matrix RF and the matrix D . Thus for the matrix product the associative law

$$(RF)D = R(FD). \quad (45.9)$$

holds.

It is easily shown that the distributive law

$$R(F + D) = RF + RD \quad (45.10)$$

is also valid.

The definition of the basic actions of matrices, given by formulae (45.3)–(45.6), corresponds completely to the analogous relations for linear operators. For formulae (45.3)–(45.5) this statement is obvious. One can easily verify this correspondence also for formula (45.6). Let the operator \hat{L} be equal to the product of the operators \hat{F} and \hat{D} , i.e. $\hat{L} = \hat{F}\hat{D}$. We find the matrix corresponding to this operator in an arbitrary representation. We

assume that the operator \hat{D} transforms the function ψ into the function φ , and the operator \hat{F} correspondingly transforms φ into χ , so that

$$\varphi = \hat{D}\psi, \quad \chi = \hat{F}\varphi. \quad (45.11)$$

We rewrite these equations, expanding the wave functions ψ , φ and χ in a series in terms of a certain system of functions ψ_m . Let the expansions

$$\psi = \sum_n c_n \psi_n; \quad \varphi = \sum_l b_l \psi_l; \quad \chi = \sum_k d_k \psi_k.$$

hold. Comparing with (44.2) and (44.5), we obtain

$$b_l = \sum_n D_{ln} c_n, \quad d_m = \sum_l F_{ml} b_l.$$

Substituting b_l into the second of these equations, we have

$$d_m = \sum_l \sum_n F_{ml} D_{ln} c_n. \quad (45.12)$$

On the other hand, taking into account that $\chi = \hat{L}\psi$, we can write that

$$d_m = \sum_n L_{mn} c_n. \quad (45.13)$$

Comparing (45.13) with (45.12), we arrive at eq. (45.6).

If a matrix F has an unequal number of columns and rows, then, by striking off a certain number of columns or rows, one obtains a square matrix with an equal number of rows and columns. From this table one can calculate the determinant $\det F$ of the matrix F .

The highest possible order of this determinant is obtained when the minimum number of rows or columns is struck off. The highest order of the non-zero determinant obtained from the matrix is called the rank of the matrix.

In certain applications the sum of the diagonal elements of the matrix, often called the trace, plays an important role. By definition

$$\text{Tr } F = \sum_n F_{nn}.$$

The matrix F is called non-singular if one can construct the inverse matrix. The inverse matrix, which we shall denote by F^{-1} , satisfies the equations

$$FF^{-1} = 1; \quad F^{-1}F = 1. \quad (45.14)$$

To find the elements of the matrix F^{-1} it is necessary to find the solution of the system of linear homogeneous equations which is obtained from the definition (45.14):

$$\sum_k (F)_{mk} (F^{-1})_{kn} = \delta_{mn}, \quad \sum_l (F^{-1})_{ml} F_{ln} = \delta_{mn} \quad (45.15)$$

for all possible values m and n .

The system of equations (45.15) can be solved only in the case where the determinant, $\det F$, of the matrix F differs from zero. (It is assumed that the matrix F is a square matrix.)

Making use of (45.14) it is easy to find the matrix which is the inverse of the product of matrices $RFD\ldots$ (if it exists)

$$(RFD\ldots)^{-1} = \ldots D^{-1} F^{-1} R^{-1}. \quad (45.16)$$

Further we define the conjugate matrix F^\dagger (or the Hermitian conjugate) of the original matrix F

$$(F^\dagger)_{mn} = (F_{nm})^*. \quad (45.17)$$

We denote the complex conjugate matrix element $(F_{nm})^*$ by F_{nm}^* .

It follows directly from the definition (45.17) that

$$(F + D)^\dagger = F^\dagger + D^\dagger. \quad (45.18)$$

It is also easy to define the matrix conjugate to the product of matrices. Thus, if $L = FD$, then

$$(L^\dagger)_{mn} = L_{nm}^* = \sum_k F_{nk}^* D_{km}^* = \sum_k (D^\dagger)_{mk} (F^\dagger)_{kn} \quad (45.19)$$

and, consequently,

$$(FD)^\dagger = D^\dagger F^\dagger. \quad (45.20)$$

In particular, if $L = kD$, where k is a number, then

$$L^\dagger = k^* D^\dagger. \quad (45.21)$$

Of course, the relation (45.20) is immediately generalized to the product of any number of matrices

$$(FDR\ldots)^\dagger = \ldots R^\dagger D^\dagger F^\dagger. \quad (45.22)$$

If the matrix F is the same as its conjugate matrix F^\dagger , $F = F^\dagger$, then it is Hermitian or self-adjoint. This definition is analogous to the definition of the hermitician property of an operator (see (44.11)). For the matrix elements of a Hermitian matrix we have

$$F_{nm} = (F^\dagger)_{nm} = F_{mn}^* . \quad (45.23)$$

The matrix F is called unitary if $F^\dagger = F^{-1}$. This condition can also be rewritten as follows:

$$F^\dagger F = FF^\dagger = 1 . \quad (45.24)$$

If the matrices F and D are unitary, then their product is also unitary. Indeed,

$$(FD)^\dagger = D^\dagger F^\dagger = D^{-1} F^{-1} = (FD)^{-1} . \quad (45.25)$$

We shall also encounter the simplest functions of matrices, where the definition of a function of the matrix means the definition of the law according to which one matrix is compared to another. Since the rules for the multiplication and addition of matrices are defined, one can easily introduce the notion of an integer rational function $f(D)$ of the matrix D . Furthermore, we shall also deal with more complex functions, for example of the form e^D , where D is a matrix. The function e^D is understood to be the following series:

$$e^D = 1 + D + \frac{1}{2} D^2 + \dots + \frac{1}{n!} D^n + \dots \quad (45.26)$$

We shall show, for example, that the matrix R , defined by the function e^{iF} , $R = e^{iF}$, where F is an arbitrary Hermitian matrix, is unitary. By a direct calculation it is easily checked that the matrix R^{-1} , the inverse of the matrix R , is the matrix e^{-iF} . On the other hand, for the matrix conjugate to R we have

$$\begin{aligned} R^\dagger &= \left(1 + iF - \frac{1}{2} F^2 - \frac{i}{3!} F^3 + \dots \right)^\dagger = \\ &= \left(1 - iF - \frac{1}{2} F^2 + \frac{i}{3!} F^3 + \dots \right) = e^{-iF} \end{aligned}$$

(because $F^\dagger = F$). Thus $R^\dagger = R^{-1}$ and, consequently, the matrix R is unitary.

The rules of action formulated in this section also remain valid for matrices with an infinitely large number of columns and rows, provided all sums (45.6), are convergent.

As we have already mentioned, the introduction of matrices is closely associated with the concept of the linear transformation of an n -dimensional vector. The concept of an n -dimensional vector is a natural generalization of

the ordinary concept of a vector. The vector \mathbf{x} in n -dimensional space is defined by the whole set of n , in general complex, numbers which are called the components of this vector x_1, x_2, \dots, x_n . Each of the components can be represented by a segment on one of the n mutually perpendicular axes in n -dimensional space. It is hardly necessary to mention that an n -dimensional space is not associated with physical reality, and that a vector in n -dimensional space is a mathematical generalization. As in the case of ordinary vectors, one can introduce the concept of the scalar product of two vectors \mathbf{x} and \mathbf{y} in n -dimensional space. Namely, the scalar product of the vectors \mathbf{y} and \mathbf{x} is defined as

$$(\mathbf{x} \cdot \mathbf{y}) = \sum_{i=1}^n x_i^* y_i. \quad (45.27)$$

If the vector \mathbf{x} has real components, then the definition (45.27) is the same as the ordinary definition of the scalar product. On the contrary, if the components of one or both of the vectors \mathbf{x} and \mathbf{y} are complex, then the definition (45.27) leads to new important consequences.

We form the scalar product of the vector \mathbf{x} and the same vector \mathbf{x}

$$(\mathbf{x} \cdot \mathbf{x}) = \sum_{i=1}^n x_i^* x_i = \sum_{i=1}^n |x_i|^2. \quad (45.28)$$

It represents a generalization of the concept of the square of a vector to the case of complex values of the components. The quantity

$$\left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \quad (45.29)$$

is said to be the length or norm of the vector. The scalar product of two vectors in n -dimensional space is, obviously, non-commuting.

$$(\mathbf{x} \cdot \mathbf{y}) = \sum_{i=1}^n x_i^* y_i \neq (\mathbf{y} \cdot \mathbf{x}) = \sum_{i=1}^n y_i^* x_i.$$

If the scalar product of two vectors is equal to zero $(\mathbf{x} \cdot \mathbf{y}) = 0$, then such vectors are called mutually orthogonal.

Let us consider two systems of coordinates k and k' each with mutually orthogonal axes. Let the components of a certain vector in the system k be x_i , and in the system k' be x'_i . As in three-dimensional Euclidean geometry,

there is a linear relation between the components expressed by

$$x'_i = \sum_k a_{ik} x_k. \quad (45.30)$$

The whole set of numbers a_{ik} forms the matrix

$$\|a\| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix},$$

which is called the linear transformation matrix.

For the linear orthogonal transformation (45.30) the following condition holds:

$$(\mathbf{x} \cdot \mathbf{x}) = \sum_{i=1}^n |x_i|^2 = (\mathbf{x}' \cdot \mathbf{x}') = \sum_{i=1}^n |x'_i|^2. \quad (45.31)$$

The transformation (45.30), satisfying the requirement (45.31), i.e. leaving the square of the length of the vector unchanged is also called a unitary transformation.

The concept of a vector with complex values of the components in an n -dimensional space can be generalized directly to the case of a space with an infinite number of dimensions, $n \rightarrow \infty$. A space with an infinite number of dimensions, for which the definition (45.28) of the square of the length of a vector is valid, is called a Hilbert space. A vector in a Hilbert space has an infinite number of components each of which can be real as well as complex.

Vectors in an n -dimensional space (with a finite as well as an infinite number of dimensions) are often written in the form of a matrix

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}.$$

Then the transformation (45.30) can be written in the form $\mathbf{x}' = \mathbf{a}\mathbf{x}$. Indeed, for the components we have $x'_i = a_{ik}x_k$, which is the same as (45.30).

§46. Geometric interpretation of the wave function. Canonical transformations

The mathematical apparatus given briefly above, in particular the vector calculus in Hilbert space, in spite of its unusual and abstract character turns out to correspond exactly to the quantum-mechanical description of the properties of microsystems. We shall consider the wave function ψ characterizing the state of the system as a vector ψ in a Hilbert space with an infinite number of dimensions. To each quantum-mechanical quantity F characterizing a property of the system there corresponds a definite system of coordinate axes or, what is the same, a system of unit basis vectors $\psi_1(x), \psi_2(x), \dots, \psi_n(x), \dots$. This system of basis vectors (basis functions) is none other than the system of eigenfunctions of the operator \hat{F} which correspond to the possible eigenvalues F_1, F_2, \dots (we assume that the spectrum is discrete; the generalization to a continuous spectrum is given farther). The components of the vector ψ in the chosen system of coordinates will be the amplitudes c_1, c_2, \dots, c_n defined by the relation

$$\psi(x) = \sum_k c_k \psi_k. \quad (46.1)$$

The amplitudes c_k , as we know (see §19), are equal to

$$c_k = \int \psi_k^*(x) \psi(x) dV.$$

On the other hand, this equality can be considered as the scalar product of the vector ψ and the vector ψ_k

$$c_k = \psi_k \cdot \psi = \int \psi_k^*(x) \psi(x) dV. \quad (46.2)$$

The definition (46.2) corresponds to (45.27). Thus if we have two vectors $\varphi(x)$ and $\psi(x)$ with components b_k and c_k respectively, then the scalar product of the vector ψ and the vector φ is equal to

$$\varphi \cdot \psi = \sum_k b_k^* c_k = \int \varphi^*(x) \psi(x) dV. \quad (46.3)$$

As we know, the whole set of amplitudes c_k represents the wave function in the F -representation. Thus the whole set of components of the vector ψ in the coordinate system whose unit vectors are the eigenfunctions ψ_n of the operator \hat{F} is the wave function in the F -representation.

The system of basis vectors ψ_1, ψ_2, \dots is a system of unit vectors which are mutually orthogonal. This follows from the condition of normalization and

orthogonality of the eigenfunctions of the operator \hat{F} (see §18)

$$\int \psi_k^*(x) \psi_m(x) dV = \psi_k \cdot \psi_m = \delta_{km} . \quad (46.4)$$

Let us now consider the transition from one representation to another. For example, the transition from the representation in which the matrix F is diagonal (F -representation) to the representation in which the matrix D is diagonal (D -representation). Geometrically this means the transition from the coordinate system formed by the basis vectors ψ_m to the coordinate system formed by the basis vectors φ_k . The functions ψ_m and φ_k are the eigenfunctions of the operators \hat{F} and \hat{D} respectively. We obtain the transformation formulae if we expand the functions φ_k in terms of the system of basis functions ψ_m (assuming that there is a discrete spectrum)

$$\varphi_k = \sum_l S_{lk} \psi_l . \quad (46.5)$$

It is obvious that

$$S_{lk} = \int \psi_l^*(x) \varphi_k(x) dV = \psi_l \cdot \varphi_k . \quad (46.6)$$

We shall call the matrix S the matrix of the transformation from one representation to another (or, correspondingly, from one coordinate system to another). We can obtain definite conclusions concerning the properties of the matrix S immediately if we take into account that the system of functions φ_k as well as the system of functions ψ_m is a system of normalized orthogonal functions. Consequently,

$$\int \varphi_m^* \varphi_k dV = \sum_{i,l} S_{im}^* S_{lk} \delta_{il} = \sum_i S_{im}^* S_{ik} = \sum_i (S^\dagger)_{mi} S_{ik} = \delta_{mk} , \quad (46.7)$$

or in matrix form

$$S^\dagger S = 1 . \quad (46.8)$$

Carrying out the inverse expansion of the functions ψ_m in terms of the functions φ_k it is easy to obtain

$$SS^\dagger = 1 . \quad (46.9)$$

It follows from eqs. (46.8) and (46.9) that the matrix S is unitary, $S^\dagger = S^{-1}$. The transformation from one representation to another, carried out by the unitary matrix S , is called a unitary or canonical transformation. Geometrically it corresponds to a 'rotation' in the Hilbert space. It is also easy to obtain the direct relation between the components of an arbitrary

vector ψ in different coordinate systems. Let

$$\psi = \sum_l c_l \psi_l = \sum_k c'_k \varphi_k.$$

Making use of (46.5), we have

$$\sum_l c_l \psi_l = \sum_{l,k} c'_k S_{lk} \psi_l.$$

Equating the expressions at equal ψ_l , we obtain

$$c_l = \sum_k S_{lk} c'_k. \quad (46.10)$$

This expression can be rewritten in the form of a matrix equation, if the whole sets of amplitudes c_l and c'_k are considered as single-column matrices c and c' . Then

$$c = S c'. \quad (46.11)$$

Multiplying from the left by S^\dagger , we obtain

$$c' = S^\dagger c. \quad (46.12)$$

Further let us find how an arbitrary matrix R is transformed in the transition to another representation. We assume that in the F -representation the following relation holds:

$$b_l = \sum_m R_{lm} c_m \quad (46.13)$$

or in the matrix form

$$b = R c. \quad (46.14)$$

In the transition to another representation the amplitudes c and b are transformed into the amplitudes c' and b' according to (46.11) and (46.12). We make use of the relation (46.11) and express c and b in eq. (46.14) in terms of c' and b' . We then obtain

$$S b' = R S c'.$$

Multiplying this equation from the left by S^\dagger , we find

$$b' = S^\dagger R S c' = R' c'.$$

Thus the matrix R' , i.e. the matrix R in the new representation, has the form

$$R' = S^\dagger R S$$

or

$$(R')_{mn} = \sum_{k,l} (S^\dagger)_{mk} R_{kl} S_{ln}. \quad (46.15)$$

Let us consider certain properties of the unitary transformation. We shall show, first of all, that if a matrix D is Hermitian in one representation, then it will also be Hermitian in another representation. Indeed, according to (46.15) and (45.22)

$$D' = S^\dagger D S,$$

$$(D')^\dagger = S^\dagger D^\dagger (S^\dagger)^\dagger = S^\dagger D^\dagger S.$$

Since $D^\dagger = D$, it turns out that $D' = D'^\dagger$. The unitary transformation also conserves the form of matrix equations. Let us, for example, consider the equations

$$F + D = R; \quad PL = T.$$

Multiplying the equations from the left by S^\dagger and from the right by S and making use of (45.9) and (46.9), we obtain

$$S^\dagger F S + S^\dagger D S = S^\dagger R S,$$

$$S^\dagger P S S^\dagger L S = S^\dagger T S.$$

Using (46.15), we rewrite these equations in the new representation

$$F' + D' = R',$$

$$P' L' = T'.$$

We see that the form of the equations has not changed. We shall show also that in a unitary transformation the trace of the matrix does not change

$$\begin{aligned} \text{Tr } F' &= \sum_n F'_{nn} = \sum_{n,l,k} (S^\dagger)_{nl} F_{lk} S_{kn} = \\ &= \sum_{l,k} F_{lk} \sum_n S_{kn} (S^\dagger)_{nl} = \sum_{l,k} F_{lk} \delta_{kl} = \text{Tr } F. \end{aligned} \quad (46.16)$$

Here we have made use of (46.9).

The unitary transformation also conserves the determinant of the matrix. Indeed, since the determinant of a product of matrices is equal to the product of the determinants $\det FDR = \det F \det D \det R$, we have

$$\det F' = \det S^\dagger \det F \det S = \det F \det (S^\dagger S) = \det F. \quad (46.17)$$

The modulus of the square of the determinant of a finite unitary matrix is equal to unity. We show this

$$|\det S|^2 = (\det S)(\det S)^*,$$

but $(\det S)^* = \det S^\dagger$, because the determinant does not change when the matrix is transposed. Consequently, we obtain

$$|\det S|^2 = \det S \det S^\dagger = \det (SS^\dagger) = 1. \quad (46.18)$$

§47. The eigenfunctions and eigenvalues of an operator given in matrix form

Let us assume that the operator \hat{D} in the F -representation is given in the form of a matrix D . We shall see how one can find the eigenfunctions and eigenvalues of this operator. The equation for the eigenfunctions and eigenvalues in the F -representation has the form

$$\sum_k D_{mk} c_k^{(n)} = D_n c_m^{(n)}. \quad (47.1)$$

Here D_n is the n th eigenvalue of the matrix D , and the whole set of amplitudes $c_1^{(n)}, c_2^{(n)}, \dots$ is the eigenfunction of the operator \hat{D} in the F -representation corresponding to the n th eigenvalue. If the eigenfunction is written in the form of a matrix with one column, $c^{(n)}$, then eq. (47.1) can be rewritten in the form

$$Dc^{(n)} = D_n c^{(n)}, \quad (47.2)$$

It is easily seen that the magnitudes of the eigenvalues do not depend on the choice of the representation. Indeed, eq. (47.1) in another representation is written, according to (46.15) and (46.11), as follows:

$$D'c'^{(n)} = D'_n c'^{(n)}, \quad (47.3)$$

where $D' = S^\dagger D S$, $c' = S^\dagger c$. Substituting these values into (47.3), we obtain

$$S^\dagger D S S^\dagger c^{(n)} = D'_n S^\dagger c^{(n)}.$$

Multiplying from the left by S , we again arrive at eq. (47.2), whence it is seen that $D'_n = D_n$.

Thus the problem of finding the eigenvalues of a matrix D reduces to finding a unitary transformation which will bring the matrix D into diagonal form. The diagonal elements of such a matrix are, as we know (see §44), its

eigenvalues. Thus if S is the required unitary transformation, then

$$S^\dagger D S = D' \quad (47.4)$$

or, multiplying on the left by S ,

$$D S = S D'.$$

Taking into account that $(D')_{mn} = D_n \delta_{mn}$, we obtain

$$\sum_k D_{mk} S_{kn} = D_n S_{mn} \quad (47.5)$$

or

$$\sum_k (D_{mk} - D_n \delta_{km}) S_{kn} = 0. \quad (47.6)$$

In these equations the matrix elements S_{kn} as well as the eigenvalues D_n are unknown. If D is a square matrix having N columns and as many rows as columns, then for each D_n we have a system of N equations (for $m = 1, 2, \dots, N$). Since the system consists of homogeneous linear equations, it has a non-trivial solution on condition that its determinant reduces to zero

$$\det \begin{vmatrix} D_{11} - D_n & D_{12} & \dots & D_{1N} \\ D_{21} & D_{22} - D_n & \dots & D_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ D_{N1} & D_{N2} & \dots & D_{NN} - D_n \end{vmatrix} = 0. \quad (47.7)$$

This is an equation of the N th power with respect to the unknown D_n . On solving it we find N roots which will be the eigenvalues of the matrix D . In particular, certain values can be equal to each other; then degeneracy occurs. All eigenvalues of a Hermitian matrix D will be real. Indeed, the matrix D' is also Hermitian (see §46) and, consequently, $(D')_{nn} = (D')_{nn}^*$ or $D_n = D_n^*$. Substituting the values D_1, D_2, \dots, D_N into the system (47.6), we determine the whole set of matrix elements S_{kn} (S_{1n}, S_{2n}, \dots) for each D_n , i.e. in the end we determine the unitary transformation matrix S . Comparing (47.5) with (47.1), we see that each column

$$\begin{pmatrix} S_{1n} \\ S_{2n} \\ \vdots \\ S_{Nn} \end{pmatrix}$$

of the matrix S is an eigenfunction of the operator \hat{D} in the F -representation corresponding to a given eigenvalue D_n . Knowing the matrix S , we can find the whole set of eigenfunctions of the operator \hat{D} in the x -representation. Indeed, if in the initial F -representation the whole set $\psi_1(x), \psi_2(x), \dots$ of eigenfunctions of the operator \hat{F} were the basis functions, then in the new representation in which the matrix D is diagonal (D -representation) the eigenfunctions $\varphi_1(x), \varphi_2(x), \dots$ of the operator \hat{D} will be the basis functions. The relation between them is given by formula (46.5) which determines the functions $\varphi_k(x)$

$$\varphi_k(x) = \sum_l S_{lk} \psi_l(x). \quad (47.8)$$

If we have two matrices F and D , then by means of one and the same unitary transformation S they can be brought into diagonal form simultaneously only in the case where they commute, i.e. $FD = DF$. Indeed, suppose that F and D are brought into diagonal form, F' and D' respectively. We form the matrix $F'D'$:

$$(F'D')_{mn} = \sum_k F'_{mk} D'_{kn} = F'_{mm} D'_{mn} \delta_{mn} = (D'F')_{mn}. \quad (47.9)$$

Consequently,

$$F'D' = D'F'.$$

Since in the unitary transformation the form of the matrix equation does not change (see §46), then in the initial representation we have $FD = DF$.

Thus we have proved that commutation of matrices is necessary in order that they may be brought simultaneously into diagonal form. It is easy to show that this condition is also sufficient.

In this section we have assumed everywhere that we have been dealing with finite matrices. However, if the number N of columns and rows tends to infinity, then the mathematical problem becomes substantially more complicated. The system (47.6) will now be a system of an infinitely large number of equations. Equation (47.7) will also be of an infinitely high power. It can be shown, however, that in this case also any Hermitian matrix may be brought into a diagonal form with real eigenvalues by means of a unitary transformation. We shall not give the proof of this statement.

§48. Continuous matrices. The Dirac notation

Up to now we have assumed that variables run over a discrete sequence of values. It is clear, however, that the preceding results must be generalized to the case of continuous variables.

It turns out that this generalization can be carried out directly. All the results obtained above remain valid if all the sums in them are replaced by the corresponding integrals. For example, formula (45.6) for the matrix element of the product of two matrices will now have the form

$$L = FD, \\ L_{\alpha\beta} = \int F_{\alpha\gamma} D_{\gamma\beta} d\gamma. \quad (48.1)$$

The integration is carried out over the entire region of variation of the corresponding variable. The unit matrix 1 is now defined by the equality

$$(1)_{\alpha\beta} = \delta(\alpha - \beta), \quad (48.2)$$

i.e. is replaced by the δ -function. As is easily seen the following relation holds: for an arbitrary matrix F

$$F1 = 1F = F.$$

The formulae of §3, which express the transformation of the wave function from the coordinate representation to the momentum representation and vice versa, can also be written in matrix form.

We note, first of all, that in its own representation the coordinate q must be expressed by a diagonal matrix (see §44). Confining ourselves for simplicity to the one-dimensional case, we have in correspondence with (48.2)

$$q_{xx'} = x\delta(x - x'). \quad (48.3)$$

Let us find the matrix representing the momentum of a particle in this representation. As in §26, we shall proceed from the relation

$$pq - qp = \frac{\hbar}{i}. \quad (48.4)$$

We shall show that this relation is satisfied if the matrix p is chosen in the form

$$p_{xx'} = \frac{\hbar}{i} \frac{\partial}{\partial x} \delta(x - x'). \quad (48.5)$$

First of all, equating the matrix elements of the left-hand and right-hand sides

of relation (48.4), we have

$$\int (p_{xx''} q_{x''x'} - q_{xx''} p_{x''x'}) dx'' = \frac{\hbar}{i} \delta(x - x').$$

Substituting (48.3) and (48.5), we obtain

$$\int \left[x'' \delta(x'' - x') \frac{\partial}{\partial x} \delta(x - x'') - x \delta(x - x'') \frac{\partial}{\partial x''} \delta(x'' - x') \right] dx'' = \delta(x - x').$$

We take the integrals in accordance with the rules of action on δ -functions (see Appendix III, Vol. 1). Since $y\delta'(y) = -\delta(y)$ (see eq. (III.8)), we find that

$$-(x - x') \frac{\partial}{\partial x} \delta(x - x') = \delta(x - x').$$

Consequently, we have proved that the matrices (48.3) and (48.5) satisfy relation (48.4). If we act, according to rules (44.5), with the matrix $p_{xx'}$ on a certain function $\psi(x)$, then we obtain the function $\varphi(x)$ equal to

$$\begin{aligned} \varphi(x) &= \int p_{xx'} \psi(x') dx' = \frac{\hbar}{i} \int \frac{\partial}{\partial x} \delta(x - x') \psi(x') dx' = \\ &= -\frac{\hbar}{i} \int \psi(x') \frac{\partial}{\partial x'} \delta(x - x') dx'. \end{aligned}$$

Integrating by parts, we obtain

$$\varphi(x) = \frac{\hbar}{i} \int \delta(x - x') \frac{\partial \psi}{\partial x'} dx' = \frac{\hbar}{i} \frac{\partial \psi}{\partial x}.$$

We see that the action of the matrix $p_{xx'}$ is equivalent to the action of the operator $\hat{p} = (\hbar/i) \partial/\partial x$. The formula of transformation of the wave function from the coordinate representation to the momentum representation is of the form

$$c(p) = (2\pi\hbar)^{-\frac{1}{2}} \int \psi(x) e^{-(i/\hbar)px} dx. \quad (48.6)$$

In matrix form this relation, according to (46.12), can be rewritten as

$$c(p) = \int S_{px}^\dagger \psi(x) dx, \quad (48.7)$$

where $S_{px}^\dagger = (2\pi\hbar)^{-\frac{1}{2}} e^{-(i/\hbar)px}$ is the unitary matrix of the transformation from the x -representation to the p -representation. It is natural that the inverse transformation is accomplished by the matrix S_{xp}

$$\psi(x) = \int S_{xp} c(p) dp, \quad (48.8)$$

where

$$S_{xp} = (2\pi\hbar)^{-\frac{1}{2}} e^{(i/\hbar)px}.$$

The form of the matrix of the coordinate q in the p -representation is easily determined by means of the matrix S_{xp} . According to (46.15), we have

$$q_p = S^\dagger q_x S$$

or

$$(q_p)_{p'p''} = \int S_{p'\tau}^\dagger(q_x)_{\tau\tau'} S_{\tau'p''} d\tau d\tau'.$$

Substituting the value of the matrices S and q_x (48.3) into the integral, we obtain

$$\begin{aligned} (q_p)_{p'p''} &= \frac{1}{2\pi\hbar} \int e^{-(i/\hbar)p'\tau} \tau \delta(\tau - \tau') e^{(i/\hbar)\tau p''} d\tau d\tau' = \\ &= \frac{1}{2\pi\hbar} \int e^{-(i/\hbar)p'\tau} \tau e^{(i/\hbar)p''\tau} d\tau = \\ &= \frac{1}{2\pi i} \frac{\partial}{\partial p''} \int_{-\infty}^{\infty} e^{(i/\hbar)(p'' - p')\tau} d\tau = \frac{\hbar}{i} \frac{\partial}{\partial p''} \delta(p'' - p'), \end{aligned}$$

so that we have obtained the matrix of the coordinate in the p -representation

$$(q_p)_{p'p''} = \frac{\hbar}{i} \frac{\partial}{\partial p''} \delta(p'' - p'). \quad (48.9)$$

Of course, this result could also be obtained directly from relation (48.4), since the matrix p is diagonal in its own representation.

Knowing the expressions for the matrices q and p , we can find the matrix of an arbitrary function of q and p . Thus, if $H(p, q)$ is a certain function of p and q , then the matrix H can be obtained if, instead of p and q , we substitute the corresponding matrices and carry out the necessary operations according to the rules of matrix addition and multiplication. Here the matrix H , as well as the corresponding operator, is understood in the sense of an expansion in a power series in terms of q and p .

Suppose that $H(p, q)$ is a known function — the Hamiltonian of a system. In the coordinate representation the matrices q and p are given by expressions (48.3) and (48.5). Consequently, the matrix H is also known in this representation. By means of a certain unitary transformation S this matrix can be transformed to the diagonal form $H' = E_n \delta_{nm}$. For definiteness we shall assume that the matrix H' has a discrete spectrum. Otherwise $\delta(n - m)$ not

δ_{nm} must be written

$$H' = S^\dagger H(p, q) S.$$

We define the matrix S . Since $HS = SH'$, we have

$$\int H_{xx'} S_{x'n} dx' = \sum_m S_{xm} H'_{mn} = E_n S_{xn} \dots \quad (48.10)$$

We consider the integrals standing on the left for different functions $H(p, q)$. First of all,

$$\begin{aligned} \int q_{xx'} S_{x'n} dx' &= x S_{xn}, \\ \int p_{xx'} S_{x'n} dx' &= \frac{\hbar}{i} \frac{\partial}{\partial x} S_{xn}. \end{aligned}$$

Further, if $U(q)$ is a certain function of q , then its matrix, as is easily seen, has the form $U_{xx'} = U(x) \delta(x - x')$, and the integral is equal to

$$\int U_{xx'} S_{x'n} dx' = U(x) S_{xn}.$$

We also obtain analogous results for functions of p . For example, the matrix of the quantity p^2 is equal, according to the rules of matrix multiplication, to

$$\int (p^2)_{xx'} = \left(\frac{\hbar}{i} \right)^2 \frac{\partial^2}{\partial x^2} \delta(x - x').$$

Correspondingly, the integral

$$\int (p^2)_{xx'} S_{x'n} dx' = \left(\frac{\hbar}{i} \right)^2 \int \frac{\partial^2}{\partial x^2} \delta(x - x') S_{x'n} dx' = \left(\frac{\hbar}{i} \right)^2 \frac{\partial^2}{\partial x^2} S_{xn}.$$

Consequently, for an arbitrary function $H(p, q)$ we obtain

$$\int H_{xx'} S_{x'n} dx' = H \left(x, \frac{\hbar}{i} \frac{\partial}{\partial x} \right) S_{xn} = \hat{H} S_{xn}. \quad (48.11)$$

Relation (48.10) is evidently none other than the Schrödinger equation written in matrix form in the x -representation:

$$\int H_{xx'} \psi_n(x') dx' = E_n \psi_n(x).$$

This equation is also easily rewritten in the p -representation in operator

form as well as in matrix form. Setting

$$\hat{H} = \frac{\hat{p}^2}{2m} + U(x)$$

and denoting the function $\psi_n(x)$ in the p -representation by $c_n(p)$ (see (48.6)), we obtain

$$\left(\frac{p^2}{2m} + \hat{U}(p) \right) c_n(p) = E_n c_n(p), \quad (48.12)$$

where $\hat{U}(p)$ is the operator of the potential energy in the p -representation.

In matrix form eq. (48.12) is

$$\frac{p^2}{2m} c_n(p) + \int U_{pp'} c_n(p') dp' = E_n c_n(p). \quad (48.12')$$

Here $U_{pp'}$ — the matrix of the operator \hat{U} — is constructed by means of (48.9) or, what is the same, is defined by

$$U_{pp'} = \int U(x) e^{-i(p-p')x/\hbar} dx.$$

Making use of (48.11), we rewrite eq. (48.10) in the form

$$\hat{H} S_{xn} = E_n S_{xn}.$$

We see that the matrix S is constructed from the eigenfunctions $\psi_n(x)$ of the operator \hat{H}

$$S_{xn} = \psi_n(x). \quad (48.13)$$

As is known, if we have a certain operator \hat{F} , then the matrix of this operator in the energy representation is given by the relation

$$F_{nm} = \int \psi_n^* \hat{F} \psi_m dx.$$

On the other hand, the same relation can be considered as the unitary transformation from the coordinate representation, in which the quantity F is defined by the matrix $F_{xx'}$, to the energy representation. The matrix of the unitary transformation is given by formula (48.13). Indeed,

$$F_{nm} = \int S_{nx}^\dagger F_{xx'} S_{x'm} dx dx'$$

and

$$\int F_{xx'} S_{x'm} dx' = \hat{F} S_{xm}$$

by analogy with (48.11). Consequently,

$$F_{nm} = \int S_{nx}^\dagger \hat{F} S_{xm} dx = \int \psi_n^*(x) \hat{F} \psi_m(x) dx.$$

and we have again arrived at the preceding relation. Thus all the relations obtained in this chapter for matrices are generalized directly to the case of operators given in differential form. Keeping this fact in mind, we shall henceforth in using the word 'operator' understand that the operator can be given in differential form as well as in matrix form.

Finally, we shall dwell briefly on a certain notation proposed by Dirac, since it is frequently encountered in the literature.

The wave function ψ or, more precisely, the set of its components in a certain coordinate system (in a certain representation) is called by Dirac a ket-vector and denoted by $|\psi\rangle$. For example, the wave function ψ_{nlm} , describing the state with given quantum numbers n, l, m , is denoted by $|nlm\rangle$. On the other hand, the complex-conjugate function is said to be a bra-vector and is denoted by $\langle\psi|$ (ψ_{nlm}^* is correspondingly denoted by $\langle nlm|$). The terms bra and ket come from the word 'bracket' $\langle \rangle$. In matrix notation the ket-vector corresponds to a column, while the bra-vector corresponds to a row. The scalar product of the bra-vector $\psi_b^* = \langle b|$ and the ket-vector $\psi_a = |a\rangle$ is denoted by $\langle b|a\rangle$, i.e.

$$\int \psi_b^*(x) \psi_a(x) dx = \langle b|a\rangle. \quad (48.14)$$

On the other hand, this scalar product can evidently be treated as the wave function ψ_a in the b -representation. Indeed, if we write the expansion

$$\psi_a(x) = \int c_a(b) \psi_b(x) db \quad (48.15)$$

(for a discrete spectrum the integral is replaced by the sum), then $c_a(b)$ represents the wave function of the state a in the b -representation

$$c_a(b) = \int \psi_b^*(x) \psi_a(x) dx = \langle b|a\rangle. \quad (48.16)$$

Correspondingly, the wave function of the state a in the x -representation $\psi_a(x)$ in the Dirac notation is of the form

$$\psi_a(x) = \langle x|a\rangle. \quad (48.17)$$

In this notation the expression (48.15) can be rewritten as

$$\langle x|a\rangle = \int_b \langle x|b\rangle \langle b|a\rangle db. \quad (48.18)$$

From (48.16) there follows the relation

$$\langle b|a\rangle = \langle a|b\rangle^*, \quad (48.19)$$

connecting the wave function of the state a in the b -representation with the wave function of the state b in the a -representation. The wave function describing a state with given momentum in the coordinate representation $\psi_p(\mathbf{r})$, in the Dirac notation has the form

$$\psi_p(\mathbf{r}) = (2\pi\hbar)^{-\frac{3}{2}} e^{(i/\hbar)\mathbf{p}\cdot\mathbf{r}} = \langle \mathbf{r}|\mathbf{p}\rangle. \quad (48.20)$$

Correspondingly we write the expansion of an arbitrary function $\psi(\mathbf{r})$ in terms of plane waves as

$$\langle \mathbf{r}|\psi\rangle = \int_p \langle \mathbf{r}|\mathbf{p}\rangle \langle \mathbf{p}|\psi\rangle d\mathbf{p} \quad (48.21)$$

or

$$\langle \mathbf{r}|\rangle = \int_p \langle \mathbf{r}|\mathbf{p}\rangle \langle \mathbf{p}|\rangle d\mathbf{p}. \quad (48.22)$$

The eigenfunction of the angular momentum operator \hat{L}^2 in the coordinate representation in the Dirac notation has the form

$$Y_{lm}(\vartheta, \varphi) = \langle \vartheta, \varphi|l, m\rangle = \langle r^{-1}\mathbf{r}|l, m\rangle. \quad (48.23)$$

The function $\langle r^{-1}\mathbf{r}|l, m\rangle$ carries out the transition from the representation lm to the coordinate representation. On the contrary, the function $Y_{lm}^*(\vartheta, \varphi) = \langle r^{-1}\mathbf{r}|l, m\rangle^* = \langle l, m|r^{-1}\mathbf{r}\rangle$ (see (48.19)) carries out the transition from the coordinate representation to the angular representation.

In the case where the angles θ, φ define the direction of the momentum vector, the function

$$Y_{lm}(\theta, \varphi) = \langle \theta, \varphi|l, m\rangle = \langle p^{-1}\mathbf{p}|l, m\rangle \quad (48.24)$$

carries out the transition from the momentum representation to the representation lm . It is the eigenfunction of the operator \hat{L}^2 in the momentum representation.

The matrix element F_{ba} in the Dirac notation is of the form

$$F_{ba} = \int \psi_b^* \hat{F} \psi_a dV = \langle b|F|a\rangle. \quad (48.25)$$

The quantities a and b , characterizing the state of the system, can run over a discrete as well as a continuous set of values. If each of the states a and b is characterized by a set of quantum numbers, for example n', l', m' and n, l, m , then the matrix element, usually denoted by $F_{n'l'm';nlm}$ or by $F_{nlm}^{n'l'm'}$, has, in the Dirac notation the form $\langle n'l'm'|F|nlm\rangle$.

§49. The Schrödinger representation, the Heisenberg representation and the interaction representation

In this section we shall discuss certain problems connected with the further development and generalization of the mathematical apparatus of quantum mechanics. We shall consider methods for the description of the development of a process in time.

Up to now we have based our considerations exclusively on the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H}\psi,$$

according to which the wave function $\psi(x, t)$ of the system could be found at an arbitrary instant of time t if its initial value $\psi(x, 0)$ was known. In this approach to the development of the process in time there corresponds a change in the wave function of the system $\psi(x, t)$.

The development of a process in time can be described by means of the operator $\hat{V}(t)$ acting on the wave function defined at a certain initial instant of time

$$\psi(x, t) = \hat{V}(t) \psi(x, 0). \quad (49.1)$$

Here we have taken as the initial time the instant $t = 0$. Of course, we could as well take as the initial time an arbitrary instant $t = t_0$. Substituting expression (49.1) into the Schrödinger equation we obtain the equation for the operator $\hat{V}(t)$

$$i\hbar \frac{\partial \hat{V}(t)}{\partial t} = \hat{H}\hat{V}(t) \quad (49.2)$$

under the condition (see (49.1)) that $\hat{V}(0) = 1$. If the operator \hat{H} does not depend explicitly on time, then the solution of eq. (49.2) can be written formally as

$$\hat{V}(t) = e^{-(i/\hbar)\hat{H}t}, \quad (49.3)$$

where the exponent is understood in the sense of an expansion in a power series.

The operator $\hat{V}(t)$ is evidently unitary

$$\hat{V}^\dagger(t) \hat{V}(t) = 1.$$

The unitary property of the operator $\hat{V}(t)$ has a simple meaning: it corresponds to the conservation in time of the normalization condition of the wave

function

$$\begin{aligned}\int \psi^*(x, 0) \psi(x, 0) dV &= \int \psi^*(x, t) \psi(x, t) dV = \\ &= \int \hat{V}^* \psi^*(x, 0) \hat{V} \psi(x, 0) dV = \int \psi^*(x, 0) \hat{V}^\dagger \hat{V} \psi(x, 0) dV.\end{aligned}$$

Thus the description of the development of a system in time amounts to the fact that the wave function or the state vector $\psi(x, t)$ changes in time. This change can be characterized by means of the unitary operator $\hat{V}(t)$ acting on the initial wave function $\psi(x, 0)$ and transforming it at every given instant to the function $\psi(x, t)$. Here the operators characterizing the system, for example the operators \hat{x} , \hat{p} or any operators $\hat{F}(\hat{x}, \hat{p})$, do not explicitly change in time.

If the state of the system is characterized by means of a Hilbert space, then the trend of development of the system can be described in the following way. Let a system of unit vectors in Hilbert space be given. This system is defined by a system of eigenfunctions of the operators forming a complete set for the given system. At the initial instant of time the state of the system is defined by the state vector $\psi(x, 0)$. The development of the system in time corresponds to a rotation of the state vector ψ in Hilbert space. Its length $(\psi \cdot \psi)$ has a constant value. Such a description of the system, in which the wave function changes in time whereas the operators are time independent, is called the Schrödinger representation. We note that the word 'representation' has in this case a more general meaning than that which it has had up to now, and characterizes a method of describing the change of a state in time. In particular, one can define the state of a system in the Schrödinger coordinate representation, in the Schrödinger momentum representation, in the Schrödinger energy representation and so on. Up to now, when speaking of the definition of the wave function in one or another representation, we have borne in mind the corresponding Schrödinger representation. The operators \hat{x} , \hat{p} and in general \hat{F} , as well as the operators of the corresponding derivatives with respect to time, $\hat{\dot{x}}$, $\hat{\dot{p}}$ and $\hat{\dot{F}}$, do not change in time in the Schrödinger representation (we assume that there are no non-stationary external fields). Indeed, according to (31.2) the operator $\hat{\dot{F}}$, for example, has the form

$$\hat{\dot{F}} = [\hat{H}, \hat{F}]$$

and does not change in time, since the operators \hat{F} and \hat{H} do not change. The matrix elements of the operator $\hat{\dot{F}}$ can also easily be determined:

$$(\hat{\dot{F}})_{mn} = [\hat{H}, \hat{F}]_{mn}. \quad (49.4)$$

In the energy representation (the Schrödinger energy representation), i.e. in a

representation such that the matrix \hat{H} is diagonal, relation (49.4) has the form

$$(\dot{F})_{mn} = \frac{i}{\hbar} (H_{mm}F_{mn} - F_{mn}H_{nn}) = i\omega_{mn}F_{mn}, \quad (49.5)$$

where

$$\omega_{mn} = \frac{1}{\hbar} (E_m - E_n),$$

$$F_{mn} = \int \psi_m^*(x) \hat{F} \psi_n(x) dV.$$

The matrix $(\dot{F})_{mn}$, as well as the matrix (F_{mn}) , does not depend explicitly on the time.

In addition to the Schrödinger representation, use is often made in quantum mechanics of another representation, called the Heisenberg representation.

In the Heisenberg representation the development of a system in time is described by means of time-dependent operators. In this case the wave function $\Phi(x)$ itself is assumed to be dependent only on coordinates, but to be time-independent. The development in the Heisenberg representation can be pictured as a rotation of the system of basis vectors in the Hilbert space with respect to the motionless state vector $\Phi(x)$.

In the general case the transition to the Heisenberg representation is carried out by means of the unitary transformation

$$\Phi(x) = \hat{V}^{-1}(t) \psi(x, t) = \psi(x, 0), \quad (49.6)$$

where $\Phi(x)$ is the wave function (the state vector) in the Heisenberg representation.

Making use of expression (49.6) and taking into account that

$$\hat{V}^{-1}(t) = \hat{V}^\dagger(t) = e^{(i/\hbar)\hat{H}t},$$

we obtain

$$\Phi(x) = \psi(x, 0) = e^{(i/\hbar)\hat{H}t} \psi(x, t). \quad (49.7)$$

In accord with the general rules (46.15) an arbitrary operator \hat{F} given in the Schrödinger representation will, in the Heisenberg representation (we denote it by \hat{F}_H), have the following form

$$\hat{F}_H = \hat{V}^\dagger(t) \hat{F} \hat{V}(t)$$

or

$$\hat{F}_H = e^{(i/\hbar)\hat{H}t} \hat{F} e^{-(i/\hbar)\hat{H}t}. \quad (49.8)$$

At the initial instant of time the expressions for the wave functions as well as for the operators are the same in the two representations. We note that the operator \hat{H} in the Heisenberg representation will be the same as in the Schrödinger representation. $\hat{H}_H = \hat{H}$. This immediately follows from formula (49.8) if it is taken into account that the operator \hat{H} commutes with all terms of the expansion of the function $e^{(i/\hbar)\hat{H}t}$. We define the matrix elements of the operator \hat{F}_H by means of the eigenfunctions of the operator \hat{H} (the Heisenberg energy representation)

$$\begin{aligned}(F_H)_{mn} &= \sum_{k,l} (e^{(i/\hbar)\hat{H}t})_{mk} F_{kl} (e^{-(i/\hbar)\hat{H}t})_{ln} = \\ &= e^{(i/\hbar)E_m t} F_{mn} e^{-(i/\hbar)E_n t} = e^{i\omega_{mn}t} F_{mn} .\end{aligned}\quad (49.9)$$

In the energy representation only diagonal matrix elements differ from zero.

If the operator \hat{F} is a function of \hat{x} and \hat{p} , then, from formula (49.8), we shall obtain the operator \hat{F} in the Heisenberg representation by using the operators \hat{x} and \hat{p} in this representation

$$\hat{F}_H = F(\hat{x}_H, \hat{p}_H) . \quad (49.10)$$

Indeed, if, for example, $\hat{F} = \hat{p}^2$, then

$$\hat{F}_H = e^{(i/\hbar)\hat{H}t} \hat{p}^2 e^{-(i/\hbar)\hat{H}t} = e^{(i/\hbar)\hat{H}t} \hat{p} e^{-(i/\hbar)\hat{H}t} e^{(i/\hbar)\hat{H}t} \hat{p} e^{-(i/\hbar)\hat{H}t} = \hat{p}_H^2 .$$

In an analogous way it is easily verified that

$$[\hat{p}_{xH}, \hat{x}_H] = [\hat{p}_x, \hat{x}] = 1 . \quad (49.11)$$

We shall obtain the equation of motion in the Heisenberg representation by differentiating (49.8)

$$\frac{\partial \hat{F}_H}{\partial t} = \frac{i}{\hbar} (\hat{H} \hat{F}_H - \hat{F}_H \hat{H}) = [\hat{H}, \hat{F}_H] . \quad (49.12)$$

If the matrix elements of the left-hand and right-hand sides of this equation are taken with the functions $\psi_n(x)$, we shall obtain, analogously to (49.5),

$$\left(\frac{\partial F_H}{\partial t} \right)_{mn} = i\omega_{mn} (F_H)_{mn} . \quad (49.13)$$

Of course, we shall arrive at exactly the same expression if we proceed from expression (49.9) and if we define the derivative of the matrix F_H with respect to time, t , as a matrix each element of which is equal to the derivative with respect to time of the corresponding matrix element of the matrix F_H ,

i.e.

$$\left(\frac{\partial F_H}{\partial t}\right)_{mn} = \frac{\partial (F_H)_{mn}}{\partial t} = i\omega_{mn}(F_H)_{mn}. \quad (49.14)$$

Thus if the operator \hat{F}_H describes a certain physical quantity, then the operator $\partial \hat{F}_H / \partial t$ correspondingly describes its derivative with respect to time.

We note that eqs. (31.7) and (31.8) can also be expressed in the Heisenberg representation. Making use of (49.11) and (49.12), we have

$$\frac{\partial \hat{x}_H}{\partial t} = \frac{1}{m} \hat{p}_H; \quad \frac{\partial \hat{p}_H}{\partial t} = -\frac{\partial \hat{U}_H}{\partial x} \quad (49.15)$$

or, in the Heisenberg energy representation,

$$\left(\frac{\partial x_H}{\partial t}\right)_{mn} = \frac{\partial}{\partial t} (x_H)_{mn} = i\omega_{mn}(x_H)_{mn} = \frac{1}{m} (p_H)_{mn}, \quad (49.16)$$

$$\left(\frac{\partial p_H}{\partial t}\right)_{mn} = \frac{\partial}{\partial t} (p_H)_{mn} = i\omega_{mn}(p_H)_{mn} = \left(-\frac{\partial U}{\partial x}\right)_{mn}. \quad (49.17)$$

The matrix relations (49.16) and (49.17) correspond in their external appearance to the classical laws of Newton.

Quantum mechanics was initially formulated by Heisenberg only as matrix mechanics. Heisenberg compared each mechanical variable to a certain matrix with elements which depend harmonically on time. The relations between matrices were taken in the form of classical relations, for example (49.17).

The Schrödinger and Heisenberg representations do not exhaust all the methods of describing quantum systems which are used in practice.

One very often has to deal in quantum mechanics with systems whose Hamiltonian can be divided into two parts: one of these, $\hat{H}^{(0)}$, represents the Hamiltonian of the system, while the other, \hat{H}' , describes the interaction of the given system with external fields or with other systems.

In this case it frequently turns out to be convenient to make use of the so-called interaction representation, introduced by Dirac.

The interaction representation is in a certain sense intermediate between the Schrödinger and Heisenberg representations. Namely, we define the wave function in the interaction representation by the relation

$$\varphi(x, t) = e^{(i/\hbar)\hat{H}_0 t} \psi(x, t) e^{-(i/\hbar)\hat{H}_0 t}. \quad (49.18)$$

Analogously, we define an arbitrary operator \hat{F} in the interaction representation as

$$\hat{F}_{\text{int}} = e^{(i/\hbar)\hat{H}_0 t} \hat{F} e^{-(i/\hbar)\hat{H}_0 t}. \quad (49.19)$$

In contrast to (49.7) and (49.8), transformation formulae (49.18) and (49.19) do not involve the total Hamiltonian but only the Hamiltonian of the system without interaction, \hat{H}_0 . The equation satisfied by the function $\varphi(x, t)$ is easily obtained. For this we differentiate relation (49.18) with respect to time and make use of the Schrödinger equation

$$\begin{aligned} i\hbar \frac{\partial \varphi(x, t)}{\partial t} &= -\hat{H}_0 \varphi(x, t) + e^{(i/\hbar)\hat{H}_0 t} (\hat{H}_0 + \hat{H}') \psi(x, t) = \\ &= e^{(i/\hbar)\hat{H}_0 t} \hat{H}' \psi(x, t) = e^{(i/\hbar)\hat{H}_0 t} \hat{H}' e^{-(i/\hbar)\hat{H}_0 t} \varphi(x, t) \end{aligned} \quad (49.20)$$

or, taking into account (49.19), we obtain

$$i\hbar \frac{\partial \varphi(x, t)}{\partial t} = \hat{H}'_{\text{int}} \varphi(x, t) \quad (49.21)$$

i.e. we have obtained the Schrödinger equation with the Hamiltonian \hat{H}'_{int} .

From relation (49.19) we find the law of change in time of an operator given in the interaction representation

$$\frac{\partial \hat{F}_{\text{int}}}{\partial t} = \frac{i}{\hbar} (\hat{H}_0 \hat{F}_{\text{int}} - \hat{F}_{\text{int}} \hat{H}_0) = [\hat{H}_0, \hat{F}_{\text{int}}] . \quad (49.22)$$

We note that the operator \hat{H}_0 has one and the same form in both the Schrödinger representation and the interaction representation.

If the operator \hat{F} depends on \hat{x} and \hat{p} , then, analogously to (49.10), it is easily shown that

$$\hat{F}_{\text{int}} = F(\hat{x}_{\text{int}}, \hat{p}_{\text{int}}) . \quad (49.23)$$

Thus we see that in the interaction representation the dependence on time of the wave function is defined by the interaction operator \hat{H}'_{int} , whereas the time-dependence associated with the operator \hat{H}_0 is directly transferred to the operators.

§50. The linear harmonic oscillator

In §10 we considered the linear harmonic oscillator by means of the Schrödinger wave equation. In fact, this problem was initially studied by the matrix method*. We shall give this solution here. On the one hand, it is a good

* M. Born, W. Heisenberg and P. Jordan, Z. Phys. 35 (1925) 557.

illustration of the use of matrix methods, and on the other hand we shall need a number of the expressions obtained in what follows.

We proceed from the known expression for the Hamiltonian of the system

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{m\omega^2 \hat{x}^2}{2}. \quad (50.1)$$

Here ω is the 'classical frequency of the oscillator'. The operators \hat{p} and \hat{x} in (50.1) are understood to be certain matrices whose mutual relation is given by eqs. (49.15). We solve the problem in the Heisenberg energy representation. Then corresponding to (49.9) we have

$$(x_H)_{nm} = x_{nm} e^{i\omega_{nm}t}. \quad (50.2)$$

The indices m, n here denote the energy levels of the system. From (49.15) and (50.1) it follows that the operator \hat{x}_H satisfies the following equation of motion:

$$\frac{\partial^2 \hat{x}_H}{\partial t^2} + \omega^2 \hat{x}_H = 0. \quad (50.3)$$

We see that the equations of motion in the Heisenberg representation have the same form as the ordinary equations of motion of classical mechanics. However, in the former the classical coordinate x is replaced by the quantum-mechanical operator \hat{x}_H . Correspondingly, for the matrix elements x_{nm} we obtain, taking account of (50.2),

$$(\omega^2 - \omega_{nm}^2) x_{nm} = 0. \quad (50.4)$$

It follows from (50.4) that only matrix elements x_{nm} for which the condition $\omega_{nm} = \pm \omega$ or $(E_n - E_m)/\hbar = \pm \omega$ is fulfilled are different from zero. We number the states in such a way that $\omega_{n,n-1} = +\omega$, and $\omega_{n,n+1} = -\omega$. Consequently,

$$x_{nm} = 0 \quad \text{for } m \neq n \pm 1 \quad \text{and} \quad x_{nm} \neq 0 \quad \text{for } m = n \pm 1. \quad (50.5)$$

The matrix elements $x_{n,n\pm 1}$ can be determined by proceeding from the commutation relations

$$\hat{p}\hat{x} - \hat{x}\hat{p} = -i\hbar.$$

We evaluate the following:

$$\sum_k (p_{nk} x_{km} - x_{nk} p_{km}) = -i\hbar \delta_{nm}.$$

According to (49.16) and (49.9), $p_{nk} = im\omega_{nk} x_{nk}$. Substituting p_{nk} into the

above expression and taking into account (50.5), we have for $m = n$

$$x_{n,n+1}x_{n+1,n} - x_{n,n-1}x_{n-1,n} = \frac{\hbar}{2m\omega}.$$

Here we have made use of the fact that $\omega_{nm} = -\omega_{mn}$, and instead of $\omega_{n,n-1}$ and $\omega_{n,n+1}$ we have substituted respectively $+\omega$ and $-\omega$. Since the matrix of the coordinate is Hermitian, then $x_{nm} = x_{mn}^*$, and we rewrite the relation obtained in the form

$$|x_{n+1,n}|^2 - |x_{n,n-1}|^2 = \frac{\hbar}{2m\omega}. \quad (50.6)$$

It is clear from (50.6) that the squares of the moduli of the matrix elements form an arithmetical progression with the difference $\hbar/2m\omega$. Since all terms of the progression are positive, it must begin with a certain positive term to which we can assign the index $n = 0$. It is obvious that then we have $x_{1,0} \neq 0$, $x_{0,-1} \equiv 0$. Consequently, from (50.6) the equation

$$|x_{1,0}|^2 = \frac{\hbar}{2m\omega}.$$

follows. Correspondingly, for an arbitrary positive integer n we find

$$|x_{n,n-1}|^2 = \frac{n\hbar}{2m\omega}. \quad (50.7)$$

From (50.7) we obtain directly

$$x_{n,n-1} = \left(\frac{n\hbar}{2m\omega}\right)^{\frac{1}{2}} e^{i\beta}; \quad x_{n-1,n} = \left(\frac{n\hbar}{2m\omega}\right)^{\frac{1}{2}} e^{-i\beta}, \quad (50.8)$$

where β is an arbitrary phase factor. Making use of the arbitrariness in the choice of β , one can set it equal to zero. Taking into account (50.2), we correspondingly obtain for the time-dependent matrix elements

$$(x_H)_{n,n-1} = \left(\frac{n\hbar}{2m\omega}\right)^{\frac{1}{2}} e^{i\omega t}; \quad (x_H)_{n-1,n} = \left(\frac{n\hbar}{2m\omega}\right)^{\frac{1}{2}} e^{-i\omega t}. \quad (50.9)$$

Let us determine the energy levels of the system, E_n , by means of the matrices (50.8). The E_n are defined as the diagonal matrix elements of the operator \hat{H} . It follows from (50.1) that

$$H_{nn} = \frac{1}{2m} \sum_k p_{nk} p_{kn} + \frac{m\omega^2}{2} \sum_k x_{nk} x_{kn}.$$

Substituting $p_{mn} = im\omega_{mn}x_{mn}$, we find

$$H_{nn} = \frac{m}{2} \left[- \sum_k \omega_{nk} \omega_{kn} x_{nk} x_{kn} + \omega^2 \sum x_{nk} x_{kn} \right].$$

Making use of the obvious equalities $\omega_{nk} = -\omega_{kn}$, $x_{nk} = x_{kn}$ and taking into account that the matrix elements x_{nm} differ from zero only when $m = n \pm 1$, we obtain

$$\begin{aligned} H_{nn} &= \frac{1}{2}m \left(\sum_k \omega_{nk}^2 x_{nk}^2 + \omega^2 \sum_k x_{nk}^2 \right) = \frac{1}{2}m \sum_k (\omega^2 + \omega_{nk}^2) x_{nk}^2 = \\ &= \frac{1}{2}m [(\omega^2 + \omega_{n,n-1}^2) x_{n,n-1}^2 + (\omega^2 + \omega_{n,n+1}^2) x_{n,n+1}^2]. \end{aligned}$$

Making use of (50.8) and $\omega_{n,n\pm 1}^2 = \omega^2$, we have finally

$$E_n = H_{nn} = m\omega^2 \left(\frac{n\hbar}{2m\omega} + \frac{(n+1)\hbar}{2m\omega} \right) = \hbar\omega(n + \frac{1}{2}), \quad (50.10)$$

which, naturally, is the same as (10.13).

Instead of the operators \hat{p} and \hat{x} , it often turns out to be convenient to introduce the operator \hat{a} and the conjugate operator \hat{a}^\dagger defined by the relations

$$\begin{aligned} \hat{a} &= \frac{1}{2}\sqrt{2}[(m\omega\hbar^{-1})^{\frac{1}{2}}\hat{x} + i(m\omega\hbar)^{-\frac{1}{2}}\hat{p}], \\ \hat{a}^\dagger &= \frac{1}{2}\sqrt{2}[(m\omega\hbar^{-1})^{\frac{1}{2}}\hat{x} - i(m\omega\hbar)^{-\frac{1}{2}}\hat{p}]. \end{aligned} \quad (50.11)$$

It follows from (50.8) that the operators \hat{a} and \hat{a}^\dagger have the following matrix elements different from zero:

$$(a^\dagger)_{n,n-1} = (a)_{n-1,n} = n^{\frac{1}{2}}, \quad (50.12)$$

while all the remaining matrix elements are equal to zero. We see, consequently, that for the operator \hat{a}^\dagger only the matrix elements corresponding to the transition $n-1 \rightarrow n$, i.e. to a transition with an increase of the quantum number n by one, are different from zero. For the operator \hat{a} the matrix elements corresponding to the transition $n \rightarrow n-1$ are different from zero. In this connection the operators \hat{a} and \hat{a}^\dagger are called the excitation annihilation operator and the excitation creation operator. From (50.12), the following commutation relation holds for the operators \hat{a}^\dagger and \hat{a} :

$$\hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a} = 1. \quad (50.13)$$

The operator \hat{H} expressed in terms of the operators \hat{a} and \hat{a}^\dagger has the form

$$\hat{H} = \frac{1}{2}\hbar\omega(\hat{a}^\dagger\hat{a} + \hat{a}\hat{a}^\dagger) \quad (50.14)$$

and, taking into account (50.12), we again arrive at (50.10). Making use of the matrix method, one can also obtain the expressions for the wave functions of the oscillator*.

§51. The matrix elements of the angular momentum operator**

In studying the properties of angular momentum in §30 of ch. 3, we proceeded directly from expressions (30.1) and (30.2) for the angular momentum operators. In the present section we shall base our exposition only on the commutation relations (30.3), (30.3'). It turns out that this statement of the problem is of a more general character. In particular, the concrete expressions for the operators, (30.1) and (30.2), cannot be used for the study of the properties of intrinsic angular momentum (spin) which will be considered in ch. 8. However, the commutation relations of the form of (30.3) also remain valid for the intrinsic angular momentum (see §60). The study of the properties of the angular momentum based on the corresponding commutation relations is conveniently carried out in matrix form. We shall denote the matrices corresponding to the projections of the angular momentum onto the x -, y - and z -axes by $\hat{J}_x, \hat{J}_y, \hat{J}_z$. The change in notation is connected with the fact that the results obtained in this section will be valid not only for the angular momentum associated with spatial motion, the orbital angular momentum $\hat{\mathbf{L}} = (\hbar/i)(\mathbf{r} \times \nabla)$, but also for the angular momentum which is not associated with spatial motion, the spin, as well as for the total angular momentum (see §62). We also introduce the matrix \hat{J}^2 corresponding to the square of the angular momentum $\hat{J}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2$. Thus we take as our basis the following commutation relations:

$$\begin{aligned}\hat{J}_x \hat{J}_y - \hat{J}_y \hat{J}_x &= i\hbar \hat{J}_z, \\ \hat{J}_y \hat{J}_z - \hat{J}_z \hat{J}_y &= i\hbar \hat{J}_x, \\ \hat{J}_z \hat{J}_x - \hat{J}_x \hat{J}_z &= i\hbar \hat{J}_y.\end{aligned}\tag{51.1}$$

First of all, we obtain the following additional commutation rules from these

* L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

** The problems touched upon in this and next sections of this chapter are considered in more detail in the books: L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965), and E.Condon and H.Shortley, *The theory of atomic spectra* (University Press, Cambridge, 1951).

relations (the proof is analogous to that presented in §30):

$$\begin{aligned}\hat{J}_x \hat{J}^2 - \hat{J}^2 \hat{J}_x &= 0, \\ \hat{J}_y \hat{J}^2 - \hat{J}^2 \hat{J}_y &= 0, \\ \hat{J}_z \hat{J}^2 - \hat{J}^2 \hat{J}_z &= 0.\end{aligned}\tag{51.2}$$

We choose the representation in which the matrices \hat{J}^2 , \hat{J}_z and \hat{H} are diagonal. Indeed, in §47 we have proved that mutually commuting matrices can simultaneously be brought into diagonal form. The commutation of a given matrix with the matrix \hat{H} expresses the law of conservation of the corresponding quantity (see §32). Hence the assumption of the commutation of the matrices \hat{J}^2 and \hat{J}_z with \hat{H} only means the fulfillment of certain conservation laws.

We number the columns and rows of the matrices considered by the indices m, j, n . The real number m defines the projection of the angular momentum onto the z -axis, $J_z = m\hbar$. The number j characterizes the value of the total angular momentum, and the number n is associated with the energy level of the system. Since all the matrices considered commute with the matrices \hat{J}^2 and \hat{H} , they are diagonal in the indices j and n . Consequently, making use of the Dirac notation, we can write the matrix elements of the matrices in which we are interested in the form

$$\begin{aligned}\langle m'j'n' | \hat{H} | mjn \rangle &= E_{jn} \delta_{jj'} \delta_{mm'} \delta_{nn'}, \\ \langle m'j'n' | \hat{J}^2 | mjn \rangle &= J_j^2 \delta_{jj'} \delta_{mm'} \delta_{nn'}, \\ \langle m'j'n' | \hat{J}_z | mjn \rangle &= m\hbar \delta_{jj'} \delta_{mm'} \delta_{nn'}, \\ \langle m'j'n' | \hat{J}_x | mjn \rangle &= (J_x)_{m'm} \delta_{jj'} \delta_{nn'}, \\ \langle m'j'n' | \hat{J}_y | mjn \rangle &= (J_y)_{m'm} \delta_{jj'} \delta_{nn'}.\end{aligned}\tag{51.3}$$

Here we have denoted the eigenvalue of the square of the angular momentum by J_j^2 . For what follows it will be convenient for us to introduce also the matrices $\hat{J}_+ = \hat{J}_x + i\hat{J}_y$ and $\hat{J}_- = \hat{J}_x - i\hat{J}_y$. It is obvious that these matrices, as well as the initial matrices \hat{J}_x and \hat{J}_y , are diagonal in the indices j and n . Taking this into account, we shall in what follows drop the indices j and n .

Our problem is the determination of the spectrum of possible values of the projection of the angular momentum onto an arbitrarily oriented axis, the establishment of the relation of these quantities to the absolute value of the angular momentum $(J_j^2)^{\frac{1}{2}}$, and finding the matrices $(J_x)_{m'm}$ and $(J_y)_{m'm}$. First of all we shall show that the spectrum of possible values of the projection of the angular momentum for a given total angular momentum is

bounded both above and below. For this we make use of the matrix relation

$$\hat{\mathbf{J}}^2 - \hat{J}_z^2 = \hat{J}_x^2 + \hat{J}_y^2.$$

Equating the diagonal matrix elements of the left-hand and right-hand sides, we obtain

$$\mathbf{J}_j^2 - m^2 \hbar^2 = \sum_k [(J_x)_{mk}(J_x)_{km} + (J_y)_{mk}(J_y)_{km}] = \sum_k |(J_x)_{mk}|^2 + |(J_y)_{mk}|^2. \quad (51.4)$$

Here we have made use of the Hermitian property of the matrices J_x and J_y . Thus the right-hand side of eq. (51.4) is undoubtedly not negative. Whence the inequality

$$m^2 \hbar^2 \leq \mathbf{J}_j^2 \quad (51.5)$$

or $-(\mathbf{J}_j^2)^{\frac{1}{2}} \leq m\hbar \leq (\mathbf{J}_j^2)^{\frac{1}{2}}$ follows. We denote the values of the quantum number m corresponding respectively to the largest and smallest possible values of the component of the angular momentum along the z -axis as m_1 and m_2 . We find the spectrum of possible values of the number m by means of the matrices \hat{J}_+ and \hat{J}_- . For this we find the commutator of these matrices with the matrix \hat{J}_z . Making use of (51.1), we obtain

$$\begin{aligned} \hat{J}_z \hat{J}_+ - \hat{J}_+ \hat{J}_z &= \hbar \hat{J}_+, \\ \hat{J}_z \hat{J}_- - \hat{J}_- \hat{J}_z &= -\hbar \hat{J}_-. \end{aligned} \quad (51.6)$$

We evaluate the first of these relations

$$(J_z J_+)_{m'm''} - (J_+ J_z)_{m'm''} = \hbar (J_+)_{m'm''}.$$

Calculating the matrix element of the derivative according to rule (45.6) and taking into account that the matrix J_z is diagonal, we find

$$\hbar(m' - m'')(J_+)_{m'm''} = \hbar(J_+)_{m'm''}. \quad (51.7)$$

It follows from eq. (51.7) that the matrix \hat{J}_+ has non-zero matrix elements $(J_+)_{m'm''}$ only under the condition that $m' - m'' = 1$, i.e. for transitions corresponding to an increase of the quantum number m by unity, $m \rightarrow m + 1$. In an analogous way it is easily shown from the second equation, (51.6), that the matrix \hat{J}_- has non-zero matrix elements only for transitions with a decrease of the quantum number m by unity, i.e. $m \rightarrow m - 1$. Thus we arrive at the conclusion that, if for a given \mathbf{J}_j^2 a certain value $m\hbar$ of the z -component of the angular momentum is possible, then the values $(m+1)\hbar$, $(m-1)\hbar$, $(m+2)\hbar$, $(m-2)\hbar$ and so on are also possible. However, we have explained before that the spectrum of possible values of the number m must be bounded:

$m_2 \leq m \leq m_1$. Setting $m'' = m_1$ in equality (51.7) and taking into account that in it m' cannot assume the value $m_1 + 1$, we see that it is fulfilled only when the matrix element $(J_+)_{m_1+1, m_1}$ reduces to zero. Consequently,

$$\langle m_1 + 1 | J_+ | m_1 \rangle = 0. \quad (51.8)$$

We have an analogous situation for the minimum possible value of the number m . The corresponding equality is fulfilled here when the matrix element $(J_-)_{m_2-1, m_2}$ reduces to zero,

$$\langle m_2 - 1 | J_- | m_2 \rangle = 0. \quad (51.9)$$

Thus the possible values of the angular momentum component are equal to $m_2\hbar$, $(m_2 + 1)\hbar$, $(m_2 + 2)\hbar$, ..., $(m_1 - 1)\hbar$, $m_1\hbar$. Here the difference $m_1 - m_2$ can only be equal to a positive integer (including zero). We shall show that the values of the numbers m_1 and m_2 determine the quantity J_f^2 . Indeed, the matrix \hat{J}^2 can be written in the form

$$\hat{J}^2 = \hat{J}_- \hat{J}_+ + J_z^2 + \hbar \hat{J}_z. \quad (51.10)$$

Taking the diagonal matrix elements of the left-hand and right-hand sides, corresponding to the transition $m_1 \rightarrow m_1$, we have

$$J_f^2 = \sum_k (J_-)_{m_1 k} (J_+)_{k m_1} + \hbar^2 m_1^2 + \hbar^2 m_1.$$

Here only $k = m_1 + 1$ is possible, but then the matrix element of J_+ (51.8) reduces to zero. Consequently,

$$J_f^2 = \hbar^2 m_1(m_1 + 1).$$

On the other hand, eq. (51.10) can also be rewritten in the form:

$$\hat{J}^2 = \hat{J}_+ \hat{J}_- + J_z^2 - \hbar \hat{J}_z. \quad (51.11)$$

If in the above expression one equates the diagonal matrix elements $m_2 \rightarrow m_2$, then

$$J_f^2 = \hbar^2 m_2(m_2 - 1)$$

and, consequently,

$$m_1(m_1 + 1) = m_2(m_2 - 1).$$

This equation is satisfied under the condition that $m_2 = m_1 + 1$ and $m_2 = -m_1$. Since, however, $m_2 \leq m_1$ always, we must retain only the second root, $m_2 = -m_1$. Consequently, the maximum (equal to $m_1\hbar$) and minimum (equal to $m_2\hbar$) possible values of the projection of the angular momentum

onto the z -axis are equal in absolute value. As we have shown, the square of the total angular momentum is equal to $\hbar^2 m_1(m_1 + 1)$. On the other hand, we decided to characterize this quantity by the quantum number j . Hence it is natural to set $m_1 = j$. In this case we have

$$\mathbf{J}_j^2 = \hbar^2 j(j + 1). \quad (51.12)$$

The possible values of the angular momentum component J_z are correspondingly equal to

$$J_z = j\hbar, (j-1)\hbar, (j-2)\hbar, \dots, (-j+1)\hbar, -j\hbar. \quad (51.13)$$

On the whole the angular momentum component assumes $2j + 1$ values. We note that, since $2j + 1$ is a positive integer, the quantum number j can take on only integer or half-integer values, $j = 0, \frac{1}{2}, 1, \frac{3}{2}$ and so on. For the orbital angular momentum this number, as we have explained in §30, takes on only integer values $j = l$. We shall see however, in ch. 8, that for intrinsic angular momentum j can also take on half-integer values.

Since the z -axis has in no way been singled out beforehand, the angular momentum projection onto any other axis is also given by formula (51.13). We note that if the number j is integer, then the angular momentum projections onto any axis are also integer (in units of \hbar); but if j is half-integer, then the angular momentum projections take on half-integer values.

Let us now find the matrices \hat{J}_x and \hat{J}_y . For this we can make use, for example, of the relation (51.10), taking the diagonal matrix elements of the left-hand and right-hand sides. Also taking into account (51.12) we have

$$\hbar^2 j(j + 1) = \sum_k (J_-)_{mk} (J_+)_{km} + \hbar^2 m^2 + \hbar^2 m,$$

where only the term of the sum with $k = m + 1$ differs from zero.

From the Hermitian property of the matrices J_x and J_y it follows that

$$(J_+)_{km} = (J_-)_{mk}^*.$$

Consequently, the preceding equation gives

$$|(J_+)_{m+1,m}|^2 = \hbar^2 [j(j + 1) - m(m + 1)] = \hbar^2 (j - m)(j + m + 1).$$

For the matrix element $(J_+)_{m+1,m}$ we have

$$\langle m + 1 | J_+ | m \rangle = \hbar [(j - m)(j + m + 1)]^{\frac{1}{2}} e^{i\beta}.$$

Without restricting the generality, the phase β can be put equal to zero. We

finally obtain

$$\begin{aligned}\langle m+1|J_+|m\rangle &= (J_x + iJ_y)_{m+1,m} = \hbar[(j-m)(j+m+1)]^{\frac{1}{2}}, \\ \langle m|J_-|m+1\rangle &= (J_x - iJ_y)_{m,m+1} = \hbar[(j-m)(j+m+1)]^{\frac{1}{2}},\end{aligned}\quad (51.14)$$

i.e.

$$(J_+)_{m+1,m} = (J_-)_{m,m+1}. \quad (51.15)$$

From the definition of the matrices J_+ and J_- it follows that

$$\hat{J}_x = \frac{1}{2}(\hat{J}_+ + \hat{J}_-), \quad \hat{J}_y = \frac{1}{2}i^{-1}(\hat{J}_+ - \hat{J}_-).$$

Making use of (51.14), we get

$$\begin{aligned}\langle m+1|J_x|m\rangle &= \langle m|J_x|m+1\rangle = \frac{1}{2}\hbar[(j-m)(j+m+1)]^{\frac{1}{2}}, \\ \langle m+1|J_y|m\rangle &= -\langle m|J_y|m+1\rangle = -\frac{1}{2}i\hbar[(j-m)(j+m+1)]^{\frac{1}{2}}.\end{aligned}\quad (51.16)$$

As an example we write the matrices which are obtained for $j = 1$:

$$\begin{aligned}J_x &= \begin{pmatrix} (J_x)_{11} & (J_x)_{10} & (J_x)_{1,-1} \\ (J_x)_{01} & (J_x)_{00} & (J_x)_{0,-1} \\ (J_x)_{-1,1} & (J_x)_{-1,0} & (J_x)_{-1,-1} \end{pmatrix} = \frac{\hbar}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \\ J_y &= \frac{\hbar}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}; \quad J_z = \hbar \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \\ J^2 &= \hbar^2 \cdot 2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.\end{aligned}\quad (51.17)$$

§52. The addition of angular momenta

We now determine the possible values of an angular momentum \mathbf{J} which is equal to the sum of two angular momenta, $\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2$. Let \mathbf{J}_1 and \mathbf{J}_2 be the angular momenta referring to two sub-systems whose mutual interaction can be disregarded. This means that the operators $\hat{\mathbf{J}}_1$ and $\hat{\mathbf{J}}_2$ act on variables referring to different sub-systems and, consequently, commute with each other,

$\hat{J}_1\hat{J}_2 = \hat{J}_2\hat{J}_1$. Since each of the operators \hat{J}_1 and \hat{J}_2 satisfies the commutation relations (51.1), the operator \hat{J} also satisfies the same commutation relations. The state of the system will be defined if the quantum numbers j_1, j_2 and m_1, m_2 characterizing the total angular momenta \mathbf{J}_1^2 and \mathbf{J}_2^2 and their projections onto an arbitrarily oriented z -axis are defined (we digress from other quantities contained in the total set, since they are not essential for what follows). For given j_1 and j_2 each of the numbers m_1 and m_2 runs respectively over $(2j_1 + 1)$ and $(2j_2 + 1)$ values. Consequently there correspond $(2j_1 + 1)(2j_2 + 1)$ states to given numbers j_1 and j_2 . However, the state of the system may be characterized by the numbers j_1, j_2, j and m , where j and m are the quantum numbers corresponding to the total angular momentum \hat{J} and its projection on the z -axis instead of by the four numbers j_1, j_2, m_1, m_2 . This means, in essence, the transition from the representation j_1, j_2, m_1, m_2 to the representation j_1, j_2, j, m . Indeed, the operators corresponding to the four latter quantities can enter into the total set just as well as the operators corresponding to the quantities j_1, j_2, m_1, m_2 . Since the obvious equality $\hat{J}_z = \hat{J}_{1z} + \hat{J}_{2z}$ holds, then the quantum number m is equal to the sum $m = m_1 + m_2$. Of course, such a simple relation does not exist for the squares of the angular momenta and we have to determine the possible values of the number j for given j_1 and j_2 . First of all we note that the maximum value of the number j is obtained if we take the largest m_1 , equal to j_1 , and the largest m_2 , equal to j_2 . Consequently, in this case $j = j_1 + j_2$. Further, we consider the following possible value of the number m : $m = j_1 + j_2 - 1$. Such a value of m can be realized either for $m_1 = j_1, m_2 = j_2 - 1$, or for $m_1 = j_1 - 1, m_2 = j_2$. Thus two independent states correspond to the given value of $m = j_1 + j_2 - 1$. Consequently, two possible values of the number j must correspond to the given m . But because the largest possible value of j is equal to $j_1 + j_2$ and because the number m cannot be larger than the number j , it is clear that to the chosen m there can correspond only $j = j_1 + j_2$ and $j = j_1 + j_2 - 1$. Choosing m one unit smaller, we obtain three states corresponding to given m :

- (1) $m_1 = j_1, m_2 = j_2 - 2$;
- (2) $m_1 = j_1 - 1, m_2 = j_2 - 1$;
- (3) $m_1 = j_1 - 2, m_2 = j_2$.

By analogy with the foregoing we arrive at the fact that the number j can assume the values $j = j_1 + j_2, j = j_1 + j_2 - 1$ and $j = j_1 + j_2 - 2$. Going on with this reasoning we find that for given j_1 and j_2 the number j can assume the values

$$j = j_1 + j_2; j_1 + j_2 - 1; j_1 + j_2 - 2, \dots, j_1 - j_2. \quad (52.1)$$

On the whole the number j assumes $2j_2 + 1$ values (under the condition that

$j_2 \leq j_1$; otherwise the indices 1 and 2 must be exchanged). The total number of states corresponding to given j_1 and j_2 is equal to

$$\sum_{j=j_1-j_2}^{j=j_1+j_2} (2j+1) = (2j_1+1)(2j_2+1),$$

as it should be. This result was obtained earlier in the so-called 'vector model' which was introduced before the appearance of quantum mechanics. In the vector model it is assumed that the length of the vector \mathbf{j} formed by adding two angular momentum vectors \mathbf{j}_1 and \mathbf{j}_2 can change by unity in a discontinuous manner. The length is maximum when these vectors are 'parallel': $\mathbf{j} = \mathbf{j}_1 + \mathbf{j}_2$, and minimum when they are 'antiparallel': $\mathbf{j} = \mathbf{j}_1 - \mathbf{j}_2$.

In the case where three or more angular momenta are to be added, we apply the rule which has been derived, adding them in pairs.

As well as the possible values of j one can find the probability that the total angular momentum of the system is equal to one or other possible value of j for given j_1 and j_2 . For this, according to general rules (§21), it is necessary to expand the wave function of the system describing the state with given values of j_1, j_2, m_1, m_2 in terms of the wave functions ψ_{jm} of the states with given j, m . Since the initial system is divided into two non-interacting sub-systems, its wave function with given j_1, m_1, j_2, m_2 can be written in the form of the product of two functions which refer respectively to each of the sub-systems $\psi_{j_1 m_1} \psi_{j_2 m_2} = \psi_{j_1 m_1}(1) \psi_{j_2 m_2}(2)$. The expansion has the form

$$\psi_{j_1 m_1} \psi_{j_2 m_2} = \sum_j C_{m_1 m_2}^j \psi_{j, m_1+m_2}. \quad (52.2)$$

The squares of the moduli of the coefficients $C_{m_1 m_2}^j$ determine the probabilities sought. We note that, since the transformation of wave functions from one representation to another is carried out by unitary matrices, we can write the expansion which is the converse of (52.2) in the form

$$\psi_{jm} = \sum_{m_2} (C_{m-m_2, m_2}^j)^* \psi_{j_1, m-m_2}(1) \psi_{j_2 m_2}(2). \quad (53.3)$$

The coefficients $C_{m_1 m_2}^j$ were calculated by Wigner by a group theory method. The reader will find a sufficiently extensive table of these coefficients in, for example, the book of Condon and Shortley†. We shall confine ourselves to the consideration of the simplest case where one of the angular

† E. Condon and H. Shortley, *The theory of atomic spectra* (University Press, Cambridge, 1951).

momenta is equal to one half, and the other is arbitrary. Thus we shall assume that $j_2 = \frac{1}{2}$. For given j_2 the number m_2 runs over only two values, namely $m_2 = \frac{1}{2}$ and $m_2 = -\frac{1}{2}$. We can correspondingly rewrite the expansion (52.3), dropping all superfluous indices, in the form

$$\psi_{jm} = C_{\frac{1}{2}} \psi_{j_1, m - \frac{1}{2}}(1) \psi_{\frac{1}{2}, \frac{1}{2}}(2) + C_{-\frac{1}{2}} \psi_{j_1, m + \frac{1}{2}}(1) \psi_{\frac{1}{2}, -\frac{1}{2}}(2). \quad (52.4)$$

We act on the left-hand and right-hand sides of this expansion with the operator $\hat{\mathbf{J}}^2 = (\hat{\mathbf{J}}_1 + \hat{\mathbf{J}}_2)^2 = \hat{\mathbf{J}}_1^2 + \hat{\mathbf{J}}_2^2 + 2\hat{\mathbf{J}}_1\hat{\mathbf{J}}_2$. The scalar product on the right is conveniently transformed so that

$$\begin{aligned} \hat{\mathbf{J}}^2 &= \hat{\mathbf{J}}_1^2 + \hat{\mathbf{J}}_2^2 + (\hat{J}_{1x} + i\hat{J}_{1y})(\hat{J}_{2x} - i\hat{J}_{2y}) + (\hat{J}_{1x} - i\hat{J}_{1y})(\hat{J}_{2x} + i\hat{J}_{2y}) + 2\hat{J}_{1z}\hat{J}_{2z} = \\ &= \hat{\mathbf{J}}_1^2 + \hat{\mathbf{J}}_2^2 + \hat{J}_{1+}\hat{J}_{2-} + \hat{J}_{1-}\hat{J}_{2+} + 2\hat{J}_{1z}\hat{J}_{2z}. \end{aligned} \quad (52.5)$$

Acting on the left-hand side of (52.4) with this operator we obtain $\hbar^2 j(j+1)$. When acting on the right-hand side it is convenient to write the operator $\hat{\mathbf{J}}^2$ in the form (52.5). In this case it should be recalled that each of the operators $\hat{\mathbf{J}}_1, \hat{\mathbf{J}}_2$ acts only on the wave function of the corresponding sub-system. It follows from the matrix relations (51.14) that

$$\begin{aligned} \hat{J}_{2-} \psi_{\frac{1}{2}, -\frac{1}{2}}(2) &= \hat{J}_{2+} \psi_{\frac{1}{2}, \frac{1}{2}}(2) = 0, \\ \hat{J}_{2-} \psi_{\frac{1}{2}, \frac{1}{2}}(2) &= \hbar \psi_{\frac{1}{2}, -\frac{1}{2}}(2), \\ \hat{J}_{2+} \psi_{\frac{1}{2}, -\frac{1}{2}}(2) &= \hbar \psi_{\frac{1}{2}, \frac{1}{2}}(2), \\ \hat{J}_{1+} \psi_{j_1, m - \frac{1}{2}}(1) &= \hbar [(j_1 - m + \frac{1}{2})(j_1 + m + \frac{1}{2})]^{\frac{1}{2}} \psi_{j_1, m + \frac{1}{2}}(1), \\ \hat{J}_{1-} \psi_{j_1, m + \frac{1}{2}}(1) &= \hbar [(j_1 - m + \frac{1}{2})(j_1 + m + \frac{1}{2})]^{\frac{1}{2}} \psi_{j_1, m - \frac{1}{2}}(1). \end{aligned}$$

Making use of these relations, we obtain the following equation:

$$\begin{aligned} j(j+1) \psi_{jm} &= [C_{\frac{1}{2}}(j_1(j_1+1) + \frac{1}{4} + m) + \\ &+ C_{-\frac{1}{2}}\{(j_1 + m + \frac{1}{2})(j_1 - m + \frac{1}{2})\}^{\frac{1}{2}}] \psi_{j_1, m + \frac{1}{2}}(1) \psi_{\frac{1}{2}, \frac{1}{2}}(2) + \\ &+ [C_{-\frac{1}{2}}(j_1(j_1+1) + \frac{1}{4} - m) + \\ &+ C_{\frac{1}{2}}\{(j_1 - m + \frac{1}{2})(j_1 + m + \frac{1}{2})\}^{\frac{1}{2}}] \psi_{j_1, m + \frac{1}{2}}(1) \psi_{\frac{1}{2}, -\frac{1}{2}}(2). \end{aligned}$$

Again substituting the expansion (52.4) for ψ_{jm} into the left-hand side and equating the coefficients of the same functions $\psi(1)\psi(2)$, we obtain two equations with respect to $C_{\frac{1}{2}}$ and $C_{-\frac{1}{2}}$. However, of these two equations only one will be independent. It gives

$$C_{-\frac{1}{2}} = C_{\frac{1}{2}} \frac{j(j+1) - j_1(j_1+1) - \frac{1}{4} - m}{[(j_1 + m + \frac{1}{2})(j_1 - m + \frac{1}{2})]^{\frac{1}{2}}}. \quad (52.6)$$

We shall obtain the second relation if we take into account that the squares of the moduli of these coefficients are equal to the corresponding probabilities

$$|C_{-\frac{1}{2}}|^2 + |C_{\frac{1}{2}}|^2 = 1. \quad (52.7)$$

The relations (52.6) and (52.7) determine the coefficients $C_{\frac{1}{2}}$ and $C_{-\frac{1}{2}}$ to within the immaterial phase factor $e^{i\alpha}$ (we choose the phase in correspondence with that taken in tables of the coefficients $C_{m_1 m_2}^j$; see the book of Condon and Shortley). Since for $j_2 = \frac{1}{2}$ the total angular momentum j can take on only two values $j_1 + \frac{1}{2}$ and $j_1 - \frac{1}{2}$, we obtain the following values of the coefficients $C_{m_1 m_2}^j$ (see table 1).

Table 1
The coefficients $C_{m_1 m_2}^j$

j	$m_2 = \frac{1}{2}$	$m_2 = -\frac{1}{2}$
$j_1 + \frac{1}{2}$	$\left(\frac{j_1 + m + \frac{1}{2}}{2j_1 + 1} \right)^{\frac{1}{2}}$	$\left(\frac{j_1 - m + \frac{1}{2}}{2j_1 + 1} \right)^{\frac{1}{2}}$
$j_1 - \frac{1}{2}$	$-\left(\frac{j_1 - m + \frac{1}{2}}{2j_1 + 1} \right)^{\frac{1}{2}}$	$\left(\frac{j_1 + m + \frac{1}{2}}{2j_1 + 1} \right)^{\frac{1}{2}}$

Perturbation Theory

§53. The theory of time-independent perturbations

The Schrödinger equation is a linear differential equation in partial derivatives with variable coefficients. Its exact solution can only be found for simple problems, some of which were considered in preceding sections.

In general the exact solution of the Schrödinger equation is associated with considerable mathematical difficulties. Hence a number of approximate methods of solving it have been devised. One such method is that of the quasi-classical approximation already considered. Another very important approximate method is the so-called perturbation theory. The term 'perturbation' and the idea of this method, which consists of a particular variant of the method of expansion in terms of a small parameter familiar in mathematics, were introduced into quantum mechanics by analogy with the perturbation method of classical mechanics. The latter played a particularly important role in solving problems of celestial mechanics.

We shall discuss perturbation theory in a general form. Its applications to the solution of actual problems will be illustrated in what follows by numerous examples.

Let us first of all consider the simplest case of a quantum-mechanical system in which the Hamiltonian operator \hat{H} does not depend explicitly on the time.

We assume that the operator \hat{H} can be written in the form

$$\hat{H} = \hat{H}_0 + \hat{H}', \quad (53.1)$$

where the operator \hat{H}' can be considered to be small in comparison with the operator \hat{H}_0 (the meaning of the word 'small' will be explained below). Then the Schrödinger equation takes the form

$$(\hat{H}_0 + \hat{H}')\psi = E\psi. \quad (53.2)$$

We further assume that the solution of the equation

$$\hat{H}_0\psi^{(0)} = E^{(0)}\psi^{(0)} \quad (53.3)$$

is known. Then for the solution of eq. (53.2) use can be made of what is, in essence, a method of successive approximation. In what follows the Hamiltonian \hat{H}_0 and the wave function $\psi^{(0)}$ will be called unperturbed, while the operator \hat{H}' will be called the perturbation operator. The 'smallness' of the operator \hat{H}' means that under the action of a perturbation the state of the system changes relatively little. Our problem is to find the solution of the Schrödinger equation assuming that the wave function $\psi^{(0)}$ of the unperturbed system is known. We shall consider the perturbations of states belonging to the discrete spectrum of the operator \hat{H}_0 . However, the operator \hat{H}_0 can have eigenvalues corresponding to a continuous spectrum as well as ones belonging to the discrete spectrum.

We seek a solution of eq. (53.2) in the form of a series in terms of the eigenfunctions of the operator \hat{H}_0

$$\psi(x) = \sum_k c_k \psi_k^{(0)}. \quad (53.4)$$

If the operator \hat{H}_0 also possesses a continuous spectrum, then we have to add the corresponding integral taken over the continuous spectrum to the sum (53.4). Substituting the sum (53.4) into eq. (53.2) and taking into account (53.3), we obtain

$$\sum_k \hat{H}' c_k \psi_k^{(0)} = \sum_k c_k (E - E_k^{(0)}) \psi_k^{(0)}.$$

We multiply the left- and right-hand sides of the equation by $\psi_m^{(0)*}$ and integrate it over the entire region of variation of the independent variables. Making use of the orthogonality of the functions $\psi_k^{(0)}$, we find

$$c_m (E - E_m^{(0)}) = \sum_k H'_{mk} c_k, \quad m = 1, 2, 3, \dots \quad (53.5)$$

where

$$H'_{mk} = \int \psi_m^{(0)*} \hat{H}' \psi_k^{(0)} dV \quad (53.6)$$

is the matrix element of the perturbation operator calculated using the unperturbed wave functions. The system of equations (53.5) is exactly equivalent to eq. (53.2). It represents the Schrödinger equation in the energy representation. We now make use of our assumption of the smallness of the perturbation operator. The energy levels and wave functions in our problem will be close to those for the unperturbed system. Hence we shall look for them in the form of the following series:

$$\begin{aligned} E &= E^{(0)} + E^{(1)} + E^{(2)} + \dots, \\ c_m &= c_m^{(0)} + c_m^{(1)} + c_m^{(2)} + \dots \end{aligned} \quad (53.7)$$

Here $E^{(0)}$ and $c_m^{(0)}$ are unperturbed values. The corrections $E^{(1)}$ and $c_m^{(1)}$ are of the same order of small quantities as the perturbation; $E^{(2)}$ and $c_m^{(2)}$ are quadratic in the perturbation and so on.

We find the correction to the n th energy level and correspondingly to the n th eigenfunction of the unperturbed system, confining ourselves to terms up to the second order of small quantities:

$$\begin{aligned} E &= E_n^{(0)} + E_n^{(1)} + E_n^{(2)}, \\ c_m &= c_m^{(0)} + c_m^{(1)} + c_m^{(2)}. \end{aligned} \quad (53.8)$$

In this section we shall assume that the n th energy level of the unperturbed system is not degenerate. For the other levels this assumption is unnecessary. In the zero order approximation the wave function is the same as the function $\psi_n^{(0)}$. This gives

$$\psi = \sum_k c_k^{(0)} \psi_k^{(0)} = \psi_n^{(0)}, \quad \text{i.e.} \quad c_k^{(0)} = \delta_{kn}. \quad (53.9)$$

Substituting (53.8) into eq. (53.5), we get

$$\begin{aligned} (\delta_{mn} + c_m^{(1)} + c_m^{(2)}) (E_n^{(0)} - E_m^{(0)} + E_n^{(1)} + E_n^{(2)}) = \\ = \sum_k H'_{mk} (\delta_{kn} + c_k^{(1)} + c_k^{(2)}). \end{aligned} \quad (53.10)$$

In eq. (53.10) one has to equate terms of the same order of small quantities. For the terms of the first order we obtain the relation

$$(E_n^{(0)} - E_m^{(0)}) c_m^{(1)} + E_n^{(1)} \delta_{mn} = \sum_k H'_{mk} \delta_{kn} = H'_{mn}. \quad (53.11)$$

Setting $m = n$, we find

$$E_n^{(1)} = H'_{nn} = \int \psi_n^{(0)*} \hat{H}' \psi_n^{(0)} dV. \quad (53.12)$$

We see that the first order correction to the energy level is equal to the mean value of the perturbation energy in the unperturbed state $\psi_n^{(0)}$. From eq. (53.11) for $m \neq n$ we find the correction of first order to the wave function

$$c_m^{(1)} = \frac{H'_{mn}}{E_n^{(0)} - E_m^{(0)}}. \quad (53.13)$$

We now write the equation for terms to the second order of small quantities:

$$(E_n^{(0)} - E_m^{(0)}) c_m^{(2)} + c_m^{(1)} E_n^{(1)} + E_n^{(2)} \delta_{mn} = \sum_k H'_{mk} c_k^{(1)}. \quad (53.14)$$

Setting $m \neq n$, we find from eq. (53.14) the correction of second order of smallness to the unperturbed wave function

$$c_m^{(2)} = \frac{1}{E_n^{(0)} - E_m^{(0)}} \left(\sum_k H'_{mk} c_k^{(1)} - E_n^{(1)} c_m^{(1)} \right). \quad (53.15)$$

The value of the amplitudes $c_n^{(1)}$ and $c_n^{(2)}$ can be obtained from the normalization condition which, taking into account (53.4), can be written in the form

$$\sum_k |c_k|^2 = 1. \quad (53.16)$$

Substituting expansion (53.8) into (53.6), we obtain

$$\sum_k |\delta_{kn} + c_k^{(1)} + c_k^{(2)}|^2 = 1. \quad (53.17)$$

We equate quantities of the same order on the left and on the right. Then we have

$$c_n^{(1)} + c_n^{(1)*} = 0, \quad c_n^{(2)} + c_n^{(2)*} + \sum_k |c_k^{(1)}|^2 = 0. \quad (53.18)$$

It follows from the relations (53.18) that the imaginary parts of the amplitude $c_n^{(1)}$ and $c_n^{(2)}$ are arbitrary quantities. The appearance of this arbitrariness is associated with the fact that the wave function is determined to within the phase factor $e^{i\alpha}$, where α can also be written in the form of a series. In corres-

pondence with this, without restricting the generality we can assume that

$$c_n^{(1)} = 0, \quad c_n^{(2)} = -\frac{1}{2} \sum_k' \frac{|H'_{kn}|^2}{(E_n^{(0)} - E_k^{(0)})^2}. \quad (53.19)$$

Here the prime on the sum indicates that the term with $k = n$ is excluded in the summation.

From (53.15) we find $c_m^{(2)}$

$$c_m^{(2)} = \sum_k' \frac{H'_{mk}H'_{kn}}{(E_n^{(0)} - E_k^{(0)})(E_n^{(0)} - E_m^{(0)})} - \frac{H'_{mn}H'_{nn}}{(E_n^{(0)} - E_m^{(0)})^2}, \quad m \neq n. \quad (53.20)$$

Setting $m = n$ in eq. (53.14), we find the second order correction to the energy level of the system:

$$E_n^{(2)} = \sum_k' \frac{H'_{nk}H'_{kn}}{E_n^{(0)} - E_k^{(0)}}. \quad (53.21)$$

The second order correction to the basic energy level turns out to be negative irrespective of the character of the perturbation. As follows from (53.8), (53.12) and (53.21), the energy of the system, to an accuracy within terms of the second order of small quantities, is equal to

$$E = E_n^{(0)} + H'_{nn} + \sum_k' \frac{|H'_{nk}|^2}{E_n^{(0)} - E_k^{(0)}}. \quad (53.22)$$

In an analogous way we obtain the expression for the perturbed wave function of the system

$$\psi = \psi_n^{(0)} + \sum_k' \frac{H'_{kn}}{E_n^{(0)} - E_k^{(0)}} \psi_k^{(0)} + \dots. \quad (53.23)$$

(We have written this formula only to within an accuracy of terms of the first order of small quantities.)

It follows from expression (53.23) that the first order correction will indeed be small if the following inequality is satisfied:

$$|H'_{kn}| \ll |E_n^{(0)} - E_k^{(0)}|. \quad (53.24)$$

Thus the perturbation theory method developed above is applicable if the matrix elements of the perturbation operator are small in comparison with the spacing between the corresponding energy levels of the unperturbed system.

§ 54. Perturbation theory in the presence of degeneracy

We now assume that the eigenvalues of the unperturbed operator \hat{H}_0 are degenerate and that the multiplicity of the degeneracy of the n th level (with energy $E_n^{(0)}$) is equal to s .

This means that the state of the unperturbed system with energy E_n is described by mutually orthogonal wave functions $\psi_{n1}^{(0)}, \dots, \psi_{ns}^{(0)}$ or by arbitrary linear combinations of them which can be chosen in such a way that the wave functions are, as before, orthogonal. When a perturbation is imposed, the eigenvalues of the operator \hat{H}_0 as a rule turn out to be non-degenerate or in any case the multiplicity of the degeneracy decreases. This fact is closely associated with the very nature of degeneracy. We have already pointed out in §35 that degeneracy is always associated with a symmetry of the Hamiltonian with respect to a definite class of transformations of the coordinates of the system. The perturbation, as a rule, does not possess the same symmetry. Hence the resulting Hamiltonian of the perturbed system will not have the previous symmetry and its energy levels will not be degenerate. Thus the perturbation removes the degeneracy. For example, in considering motion in a centrally symmetric field we have seen that the $(2l+1)$ -fold degeneracy of the energy levels is associated with the symmetry (invariance) of the Hamiltonian with respect to rotation of the system about the centre of force. If the system is now placed in an external field, then the total Hamiltonian will no longer possess spherical symmetry. The perturbation (in the given case, the external field) removes the $(2l+1)$ -fold degeneracy corresponding to the components of the angular momentum.

On imposing the perturbation the degenerate energy level splits into s close levels. To each of these energy levels there corresponds a wave function which is a linear combination of the functions $\psi_{nr}^{(0)}$

$$\psi = \sum_{m,r} c_{mr} \psi_{mr}^{(0)}. \quad (54.1)$$

As before, we consider the perturbation to be small and, in the first approximation of perturbation theory, we seek the nearby energy levels (they are often called sub-levels) into which the degenerate level splits. At the same time we seek the corresponding set of wave functions in the zero order approximation. That is, we have to find, in the zero order approximation, correct expressions for the amplitudes c_{mr} in the sum (54.1) such that the linear combination (54.1) will correspond to one of the sub-levels into which the initial energy level splits and that it will undergo only a small change when the perturbation is taken into account in the next approximation.

Let us first consider the case of two close levels. In this case formula (54.1) gives $\psi = c_1 \psi_1 + c_2 \psi_2$. Substituting this value into the Schrödinger equation (53.2), we find

$$-c_1(E - E^{(0)}) + H_{12}c_2 + H_{11}c_1 = 0,$$

$$c_1H_{21} - c_2(E - E^{(0)}) + c_2H_{22} = 0.$$

Setting $E = E^{(0)} + E^{(1)}$, we obtain the system of homogeneous equations

$$(H_{11} - E^{(1)})c_1 + H_{12}c_2 = 0,$$

$$H_{21}c_1 + (H_{22} - E^{(1)})c_2 = 0.$$

The condition for this system to have a solution is that the determinant of the coefficients be equal to zero,

$$\begin{vmatrix} H_{11} - E^{(1)} & H_{12} \\ H_{21} & H_{22} - E^{(1)} \end{vmatrix} = 0.$$

Hence

$$(H_{11} - E^{(1)})(H_{22} - E^{(1)}) = |H_{12}|^2,$$

or

$$E_{1,2}^{(1)} = \frac{1}{2}(H_{11} + H_{22}) \pm \frac{1}{2}[(H_{11} - H_{22})^2 + 4|H_{12}|^2]^{\frac{1}{2}}.$$

We see that the degenerate level splits into two levels corresponding to the two different signs in front of the square root.

If the perturbation is small, so that $|H_{12}|^2 \ll (H_{11} - H_{22})^2$, then we come back to the case of two independent levels whose energies are equal to

$$E_1 = E^{(0)} + E^{(1)} = E^{(0)} + H_{11} + \frac{|H_{12}|^2}{H_{11} - H_{22}}$$

$$E_2 = E^{(0)} + H_{22} - \frac{|H_{12}|^2}{H_{11} - H_{22}}.$$

But if the levels are lying so close to each other that $|H_{12}|^2 \gg (H_{11} - H_{22})^2$,

then we obtain

$$E_1 = E^{(0)} + \frac{1}{2}(H_{11} + H_{22}) + |H_{12}|^2 + \frac{(H_{11} - H_{22})^2}{8|H_{12}|},$$

$$E_2 = E^{(0)} + \frac{1}{2}(H_{11} + H_{22}) - |H_{12}|^2 - \frac{(H_{11} - H_{22})^2}{8|H_{12}|}.$$

Analogous results are also obtained in the general case of s -fold degeneracy.

Substituting (54.1) into the Schrödinger equation (53.2) we obtain analogously to (53.5)

$$c_{mp}(E - E_n^{(0)}) = \sum_{k,r} H'_{mp;kr} c_{kr}, \quad (54.2)$$

where

$$H'_{mp;kr} = \int \psi_{mp}^{(0)*} \hat{H}' \psi_{kr}^{(0)} dV.$$

If we are interested in perturbation of the energy level E_n , then we have to put $m = n$ and to equate terms of the first order of magnitude. But in the zero order approximation the wave function ψ is the superposition of the functions $\psi_{nr}^{(0)}$, i.e. the $c_{kr}^{(0)}$ are different from zero only for $k = n$. Writing the energy E in eq. (54.2) in the form $E = E_n^{(0)} + E^{(1)}$, we get

$$c_p^{(0)} E^{(1)} = \sum_{r=1}^s H'_{pr} c_r^{(0)} \quad (54.3)$$

(we have dropped the fixed index n in the notation).

The system of homogeneous equations (54.3) has a non-trivial solution only in the case where the determinant of the coefficients of the unknown quantities is equal to zero, i.e. under the condition

$$\begin{vmatrix} H'_{11} - E^{(1)} & H'_{12} & \dots & H'_{1s} \\ H'_{21} & H'_{22} - E^{(1)} & \dots & H'_{2s} \\ \dots & \dots & \dots & \dots \\ H'_{s1} & \dots & \dots & H'_{ss} - E^{(1)} \end{vmatrix} = 0. \quad (54.4)$$

This equation is called the secular equation. The secular equation is an equation of the s th order with respect to $E^{(1)}$ and has, consequently, s roots. Solving it with respect to $E^{(1)}$, we find s values for this quantity. This means that the n th energy level splits into s sublevels $E_n^{(0)} + E_1^{(1)}$, $E_n^{(0)} + E_2^{(1)}$, ..., $E_n^{(0)} + E_s^{(1)}$. In particular cases certain roots of the secular equation turn out

to be equal to one another. In this case the perturbation only partially removes the degeneracy in the system.

Substituting the values $E^{(1)}$ found into eq. (54.3), we can determine the amplitudes $c_{nr}^{(0)}$ corresponding to a given correction to the energy $E^{(1)}$. By this means we find, in the zeroth approximation, the correct wave functions corresponding to the energy sub-levels into which the level $E_n^{(0)}$ splits. These wave functions are slightly distorted under the action of the perturbation.

The method discussed is also applicable in the case where the eigenvalues of the operator \hat{H}_0 are not degenerate but are so closely spaced that the inequality (53.24) is not satisfied*.

As an example of the application of the methods discussed in this and the preceding sections we shall consider the displacement of the lowest energy level of the hydrogen-like atom, and the splitting of the first excited level, caused by the finite size of the nucleus.

In considering hydrogen-like atoms we assumed that the electron was in the Coulomb field of the nucleus. However, the difference between the correct field and a Coulomb field in the region of the nucleus itself was not taken into account. We now assume the nucleus to be a uniformly charged sphere of radius r_0 . Then the potential energy of the electron for $r \leq r_0$ has the form

$$U(r) = -\frac{Ze^2}{r_0} \left(\frac{3}{2} - \frac{r^2}{2r_0^2} \right). \quad (54.5)$$

The difference between the potential energy of the electron and its value for a pure Coulomb field is the perturbation Hamiltonian

$$\hat{H}' = \begin{cases} -\frac{Ze^2}{r_0} \left(\frac{3}{2} - \frac{r^2}{2r_0^2} \right) + \frac{Ze^2}{r}, & r \leq r_0 \\ 0 & r > r_0. \end{cases} \quad (54.6)$$

We define the correction to the ground energy level in the first approximation:

$$E^{(1)} = H'_{00} = \int \psi_0^* \hat{H}' \psi_0 dV. \quad (54.7)$$

The wave function of the ground state (38.22) is

$$\psi_0 = 2(Z/a)^{3/2} e^{-Zr/a} (4\pi)^{-1/2}. \quad (54.8)$$

* For more details see V.A.Fok, *Nachala kvantovoi mekhaniki (Principles of quantum mechanics)* (KUBUCH, 1932) p. 92.

Substituting (54.8) into (54.7), we obtain

$$E^{(1)} = \frac{Z^3}{a^3} 4 \int_0^{r_0} e^{-2Zr/a} \left[\frac{Ze^2}{r} - \frac{Ze^2}{r_0} \left(\frac{3}{2} - \frac{r^2}{2r_0^2} \right) \right] r^2 dr. \quad (54.9)$$

Since the radius of the first Bohr orbit is $a \sim 10^{-8}$ cm and $r_0 = 10^{-12}$ cm, the exponent of the exponential in (54.9) is very small, and the exponential can be replaced by unity. Integrating the integral in (54.9), we find

$$E^{(1)} = \frac{2Z^4 e^2}{5a} \left(\frac{r_0}{a} \right)^2 = -\frac{4}{5} E_1^{(0)} \left(\frac{r_0}{a} \right)^2 Z^2. \quad (54.10)$$

Even for the heaviest atoms, $Z \sim 100$, and the ratio $E^{(1)}/E_1^{(0)} \sim 10^{-4}$.

Let us now consider the first excited level $n = 2$. As we have shown in §38 this level will be 4-fold degenerate (the states $\psi_{200}, \psi_{211}, \psi_{210}, \psi_{21-1}$). We shall number these wave functions by the index $s = 1, 2, 3, 4$ respectively. It is clear already from general considerations that the perturbation will partially remove the degeneracy. Indeed, in the Coulomb field we have degeneracy in the two quantum numbers l and m . The degeneracy in the quantum number l is specific for the Coulomb field. However, the degeneracy in the magnetic quantum number m occurs in an arbitrary centrally symmetric field. In view of the fact that when the perturbation is taken into account the field is no longer strictly a Coulomb field, although it will remain a central field, the degeneracy in the quantum number l is removed.

Thus we can expect the level with $n = 2$ to split into 2 levels, with $n = 2, l = 0$ and $n = 2, l = 1$. We shall show that a calculation does indeed lead to this splitting.

The secular equation in this case will have the form

$$\begin{vmatrix} H'_{11} - E^{(1)} & H'_{12} & H'_{13} & H'_{14} \\ H'_{21} & H'_{22} - E^{(1)} & H'_{23} & H'_{24} \\ H'_{31} & H'_{32} & H'_{33} - E^{(1)} & H'_{34} \\ H'_{41} & H'_{42} & H'_{43} & H'_{44} - E^{(1)} \end{vmatrix} = 0. \quad (54.11)$$

The matrix elements are taken with respect to the functions ψ_{nlm} : $\psi_1 = \psi_{200}$, $\psi_2 = \psi_{211}$, $\psi_3 = \psi_{210}$ and $\psi_4 = \psi_{21-1}$.

In view of the fact that the perturbation operator \hat{H}' (54.6) depends only on the coordinate r , all non-diagonal matrix elements in (54.11) reduce to zero because of the orthogonality of the spherical functions (30.18). Indeed,

integrating with respect to the angular variables, we obtain

$$\int Y_{l'm'}^* Y_{lm} \sin \vartheta \, d\vartheta \, d\varphi = \delta_{ll'} \delta_{mm'}.$$

Making use of (38.22)–(38.24), we obtain for the diagonal matrix elements (since the integral with respect to angular variables is equal to unity):

$$H'_{11} = \frac{Z^3}{2a^3} \int_0^{r_0} \left(1 - \frac{Zr}{2a}\right)^2 e^{-Zr/a} \left(\frac{Ze^2}{r} - \frac{3Ze^2}{2r_0} + \frac{Ze^2 r^2}{2r_0^3}\right) r^2 dr, \quad (54.12)$$

$$H'_{22} = H'_{33} = H'_{44} = \frac{Z^3}{24a^3} \int_0^{r_0} e^{-Zr/a} \frac{Z^2 r^2}{a^2} \left(\frac{Ze^2}{r} - \frac{3Ze^2}{2r_0} + \frac{Ze^2 r^2}{2r_0^3}\right) r^2 dr. \quad (54.13)$$

Neglecting terms of the order r_0/a in comparison with unity, we get

$$E_1^{(1)} = \frac{1}{20} \frac{Z^4 e^2}{a} \left(\frac{r_0}{a}\right)^2, \quad (54.14)$$

$$E_2^{(1)} = \frac{1}{1120} \frac{Z^2 e^2}{a} \left(\frac{Zr_0}{a}\right)^4. \quad (54.15)$$

We see that the original level is split into two sub-levels. The displacement of each of them with respect to the position of the original level is given by formulae (54.14) and (54.15). The value of the shift of the level $n = 2, l = 0$ is smaller by about an order of magnitude than the shift of the level $n = 1, l = 0$. The shift of the level $n = 2, l = 1$ is even smaller owing to the small factor $10^{-3}(Zr_0/a)$. This is due to the fact that the electron in the state $n = 2, l = 1$ is, in the main, outside the region of the nucleus and that the distortion of the Coulomb field in this region affects its state very little.

Finally, we note that the corrections considered here turn out to be considerably more substantial for mesic atoms. This is associated with the fact that mesons are much heavier than electrons and hence are in the main much more near the nucleus (see §38). Thus for the μ -mesic atom the relative shift of the level with $l = 0$ is larger by a factor of about 4×10^4 than for the ordinary atom, and becomes an appreciable quantity.

§55. The theory of time-dependent perturbations

Perturbations acting on a quantum-mechanical system very often have a nonstationary character (i.e. depend on time). This means that the perturbation operator \hat{H}' is an explicit function of time $\hat{H}'(t)$. Numerous examples of

such perturbations will be given below. We assume that the stationary states of the unperturbed system are known, i.e. that the wave functions $\psi_n^{(0)}(x, t) = \psi_n^{(0)}(x) \exp[-(i/\hbar)E_n t]$ the unperturbed equation

$$i\hbar \frac{\partial \psi_n^{(0)}(x, t)}{\partial t} = \hat{H}_0 \psi_n^{(0)}(x, t). \quad (55.1)$$

are known. We restrict ourselves first to the simple case where the states of the unperturbed system belong to a discrete spectrum.

If the system is acted upon by a small perturbation described by the operator $\hat{H}'(t)$, then the wave function of the perturbed system ψ satisfies the equation

$$i\hbar \frac{\partial \psi}{\partial t} = (\hat{H}_0 + \hat{H}') \psi. \quad (55.2)$$

The method of approximate solution of this equation was worked out by Dirac and is often called the Dirac perturbation theory or the method of variation of constants. The state of a perturbed system depends on time and its energy is not a constant of the motion. Our problem now is not to find the stationary states of the perturbed system, because they do not exist, but the calculation of the time-dependent wave function of the system. Hence the perturbation theory method must be modified. The solution of eq. (55.2) in the method of variation of constants is written in the form of an expansion in terms of the eigenfunctions of the unperturbed problem

$$\psi(x, t) = \sum_k c_k(t) \psi_k^{(0)}(x, t). \quad (55.3)$$

Since the wave functions $\psi_k^{(0)}(x, t)$ form a complete system of functions, such an expansion is always possible. The coefficients $c_k(t)$ of the expansion are functions of time only and not of the coordinates. Substituting the expansion (55.3) into eq. (55.2), we obtain

$$i\hbar \sum_k \left(\frac{dc_k}{dt} \psi_k^{(0)}(x, t) + c_k \frac{d\psi_k^{(0)}(x, t)}{dt} \right) = \sum_k c_k (\hat{H}_0 + \hat{H}') \psi_k^{(0)}(x, t). \quad (55.4)$$

We multiply eq. (55.4) from the left by $\psi_m^{(0)*}(x, t)$ and integrate over all space. Then, taking into account eq. (55.1) and the orthogonality of the wave functions of the unperturbed system $\psi_k^{(0)}(x, t)$, we have

$$i\hbar \frac{dc_m}{dt} = \sum_k H'_{mk} \exp(i\omega_{mk} t) c_k, \quad (55.5)$$

where H'_{mk} is the matrix element of the perturbation operator

$$H'_{mk} = \int \psi_m^{(0)*}(x) \hat{H}' \psi_k^{(0)}(x) dV, \quad (55.6)$$

and

$$\omega_{mk} = \hbar^{-1}(E_m - E_k).$$

The system of equations (55.5) is exact. It is equivalent to the initial equation (55.2), since the whole set of coefficients c_k completely determines the wave function ψ . However, it is clear that the solution of the infinite system of equations (55.5) is no simpler a problem than the solution of the initial equation (55.2). Hence for a simplification of the system of equations (55.5) we have to make use of the fact that the perturbation acting on the system is small. We assume initially that for $t \leq 0$ the system was in a state with the wave function $\psi_n^{(0)}$. Then for $t \leq 0$ all coefficients in the expansion (55.3), with the exception of the coefficients with the index n , are equal to zero, i.e.

$$c_k(0) = \delta_{kn}. \quad (55.7)$$

Beginning with time $t = 0$ the system undergoes the action of a small perturbation. We assume that owing to the weakness of the perturbation the wave function of the initial state, $\psi_n(0)$, changes little with time. Correspondingly, we seek the coefficients $c_k(t)$ at an instant of time $t > 0$ in the form

$$c_k(t) = c_k^{(0)}(t) + c_k^{(1)}(t) + c_k^{(2)}(t) + \dots \quad (55.8)$$

where

$$c_k^{(0)}(t) = c_k(0) = \delta_{nk}.$$

The correction $c_k^{(1)}(t)$ is of the same order of small quantities as the perturbation, $c_k^{(2)}(t)$ is quadratic in the perturbation and so on. Substituting the expansion (55.8) into eq. (55.5), we find

$$i\hbar \frac{dc_m^{(1)}}{dt} = \sum_k H'_{mk} \exp(i\omega_{mk}t) c_k^{(0)} = H'_{mn} \exp(i\omega_{mn}t). \quad (55.9)$$

Here all terms of the second and higher order of small quantities in the perturbation have been dropped. Integrating (55.9), we get

$$c_m^{(1)}(t) = \frac{1}{i\hbar} \int_0^t H'_{mn} \exp(i\omega_{mn}t) dt. \quad (55.10)$$

In an analogous way one can find the corrections to $c_m^{(0)}$ of the second and

higher orders of small quantities. For example, for the correction of the second order $c_m^{(2)}$ one easily obtains the expression

$$c_m^{(2)} = \frac{1}{i\hbar} \sum_k \int_0^t H'_{mk} \exp(i\omega_{mk}t) c_k^{(1)} dt. \quad (55.11)$$

If the perturbation is sufficiently small, then one can restrict oneself to a small number of terms in the expansion. Thus, in principle, the wave function at any instant of time $t > 0$ can be found with the desired degree of accuracy.

§56. The transition of a system into new states under the action of perturbations

We have found that if a system in a definite energy state for $t \leq 0$ described by the wave function $\psi_n^{(0)}$ is acted upon by a perturbation $\hat{H}'(t)$, then for $t > 0$ the system turns out to be in a new state with the wave function (55.3). This means that for $t > 0$ the system can be found in any of its possible stationary quantum states; the probability of finding the system in a certain quantum state m is defined according to the general rules of quantum mechanics by the value of the quantity $|c_m|^2$. Since at the initial instant $t = 0$ the system was in the n th stationary state, then, consequently, $|c_m(t)|^2$ defines the probability of the transition of the system from the n th state into the m th state in time t , $W_{mn}(t) = |c_m(t)|^2 \equiv |c_{mn}(t)|^2$. Here we have denoted the initial state of the system by the second subscript.

Thus the perturbation turns out to give rise to the transition of the system from one quantum state into another. A characteristic property of this process, which has no analogy in classical physics, is the fact that a given perturbation gives rise to the transition of the system from a stationary state with definite energy into a new state in which the energy has no clearly defined value. This is often understood in such a way that under the action of the perturbation the system goes over by a discontinuous process into one of the possible energy states. The state into which the system goes will be a matter of chance. However, such an idea is incorrect and contradicts the physical basis of quantum mechanics. As a matter of fact, the final state is described by a wave function ψ and is hence a definite state (in the quantum-mechanical sense).

The transition from the initial to the final state is not carried out discontinuously, but proceeds in time. Indeed, as we shall see below, the transition

probability is determined by the character of the perturbation and by its dependence on time.

Transitions from a discrete into a continuous spectrum are of the greatest interest, and we shall consider such transitions in what follows.

To determine the transition probability it is evidently necessary to know the dependence of the matrix element of the perturbation operator \hat{H}'_{vn} on time. Here we characterize a state in the continuous spectrum by the index v .

Let us consider, first of all, the important case where the perturbation operator is a harmonic function of time. Then the matrix element of the perturbation operator (taken with respect to time-independent unperturbed wave functions) is also a harmonic function of time, i.e.

$$H'_{vn}(t) = H'_{vn}(0) \cos \omega t. \quad (56.1)$$

We shall assume that the frequency ω satisfies the relation

$$h\omega > E_0 - E_n^{(0)},$$

where E_0 denotes that energy value of the system with which the continuous spectrum begins. Substituting (56.1) into (55.10), we find

$$c_{vn}^{(1)} = -\frac{1}{2\hbar} H'_{vn}(0) \left[\frac{\exp[i(\omega_{vn} + \omega)t] - 1}{\omega_{vn} + \omega} + \frac{\exp[i(\omega_{vn} - \omega)t] - 1}{\omega_{vn} - \omega} \right]. \quad (56.2)$$

Here we denote the initial state of the system by the second index in $c_{vn}^{(1)}$. Since the continuous spectrum lies in a range of energies higher than the discrete spectrum, then $\omega_{vn} > 0$. From the structure of expression (56.2) it follows that for $\omega_{vn} \approx \omega$ the denominator of one of the terms is close to zero. Transitions into states for which the condition $\omega_{vn} \approx \omega$ is fulfilled occur with a low probability. From what follows it will be seen that the probability of the transition into such states increases linearly with time. Hence, dropping the first term in formula (56.2), we have

$$c_{vn}^{(1)} = -\frac{1}{2\hbar} H'_{vn}(0) \frac{\exp[i(\omega_{vn} - \omega)t] - 1}{\omega_{vn} - \omega}. \quad (56.3)$$

Correspondingly for the square of the modulus $|c_{vn}^{(1)}|^2$ we obtain

$$|c_{vn}^{(1)}|^2 = \frac{1}{4\hbar^2} |H'_{vn}(0)|^2 \frac{\sin^2 \frac{1}{2}(\omega_{vn} - \omega)t}{\frac{1}{4}(\omega_{vn} - \omega)^2} = \frac{\pi |H'_{vn}(0)|^2}{4\hbar^2} t f(\alpha, t), \quad (56.4)$$

where

$$\alpha = \frac{1}{2}(\omega_{vn} - \omega) \quad \text{and} \quad f(\alpha, t) = \frac{\sin^2 \alpha t}{\pi \alpha^2 t}.$$

Usually in practice it is of interest to know the magnitude of $|c_{vn}^{(1)}|^2$ for large values of time t (we recall that the instant of switching on the perturbation is taken as the zero of time $t = 0$). Therefore it is necessary to consider the behaviour of the function $f(\alpha, t)$ when $t \rightarrow \infty$. It is easily seen that for $\alpha \neq 0$ and $t \rightarrow \infty$, $f(\alpha, t) \rightarrow 0$. For $\alpha = 0$, $f(0, t) = t/\pi$ and increases indefinitely with increasing time t . Finally, integrating $f(\alpha, t)$ over all values of α , we find

$$\int_{-\infty}^{\infty} \frac{\sin^2 \alpha t}{\pi \alpha^2 t} d\alpha = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin^2 x}{x^2} dx = 1. \quad (56.5)$$

Comparing the above properties of the function $f(\alpha, t)$ with the properties of the δ -function, we see that they are identical (see Appendix III). Thus

$\lim_{t \rightarrow \infty} f(\alpha, t)$ is one of the possible concrete forms of the δ -function, and we can write

$$\lim_{t \rightarrow \infty} \frac{\sin^2 \alpha t}{\pi \alpha t} = \delta(\alpha) = \delta\left(\frac{\omega_{vn} - \omega}{2}\right).$$

Substituting this expression into formula (56.4) and making use of the known properties of the δ -function (see Appendix III, Vol. 1), we obtain

$$\begin{aligned} |c_{vn}^{(1)}|^2 &= \frac{\pi}{4\hbar^2} |H'_{vn}(0)|^2 t \delta\left(\frac{\omega_{vn} - \omega}{2}\right) = \\ &= \frac{\pi}{2\hbar} |H'_{vn}(0)|^2 t \delta(E_v - E_n^{(0)} - \hbar\omega). \end{aligned} \quad (56.6)$$

Formulae (56.4) and (56.6) will be valid only under the condition that the probability of the transition from the given n th state into any v th state is low, i.e.

$$\int |c_{vn}^{(1)}|^2 dv \ll 1.$$

Only in this case is the initial assumption of the smallness of change of the wave function of the initial state fulfilled. Since the transition probability increases linearly with time, it is necessary, for perturbation theory to be applicable, that the time of action of the perturbation t be not too large. Therefore we shall find out what conclusions can be drawn concerning the probability of a transition in a finite time interval t . For this we study formula (56.4) without passing to the limit $t \rightarrow \infty$, i.e. we consider the behaviour of the function $f(\alpha, t)$. The plot of this function with respect to time is shown in fig. V.17.

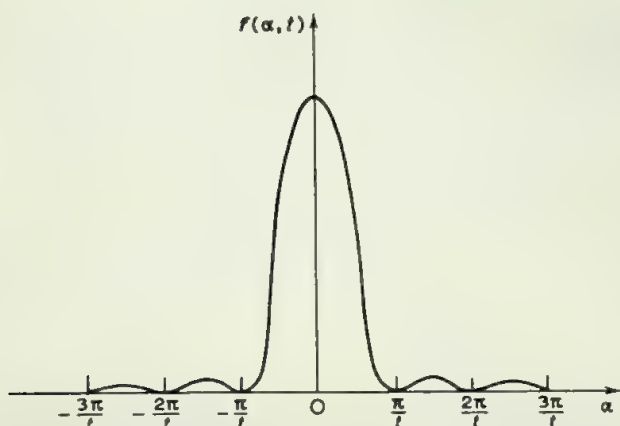


Fig. V.17

From the form of the function $f(\alpha, t)$ it follows that, in the main, transitions are realized into those states for which the quantity α lies within the limits of the principal maximum, i.e. $\Delta\alpha \sim t^{-1}$. The straggling of the values of the parameter α determines the straggling of the energy values of the final state of the system

$$\Delta E_v \sim \hbar \Delta\alpha \sim \hbar/t. \quad (56.7)$$

Thus we arrive at the conclusion that in time t the system can, under the action of perturbation (56.1), make transitions into states with energy $E_v = E_n^{(0)} + \hbar\omega + \Delta E_v$, where $\Delta E_v \sim \hbar/t$.

The uncertainty in the energy of the final state is $\Delta E_v \rightarrow 0$ as $t \rightarrow \infty$. We note that, proceeding from the uncertainty relation for time and energy (see §34), just such a value of the uncertainty in the energy of the final state $\Delta E_v \sim \hbar/t$ was to be expected.

From the requirement that the uncertainty in the energy of the final state ΔE_v be small in comparison with the energy $\hbar\omega$ the following inequality arises: $t \gg \omega^{-1}$. Consequently, $\Delta E_v \ll \hbar\omega$ if the time of action of the perturbation is large as compared to the period of perturbation.

The transition to the δ -function in formula (56.6) means that the time t must be sufficiently large for the uncertainty in the energy of the final state to be disregarded, but that nevertheless the condition of applicability of perturbation theory is still fulfilled.

Formula (56.6), containing the δ -function, has, of course, a meaning only

because integration with respect to the argument of the δ -function is subsequently implied.

We note that the conditions of applicability of perturbation theory are violated in considering transitions in a discrete spectrum in the so-called resonance case, i.e. when $|\omega_{kn}| \approx \omega$. Under these conditions the corrections to the wave function $\psi_n^{(0)}$ become large and the problem must be solved precisely*.

The probability of the transition per unit time from a quantum state with energy $E_n^{(0)}$ into a state of the continuous spectrum in the interval $d\nu$ is defined by the formula

$$dW_{\nu n} = \frac{1}{t} |c_{\nu n}^{(1)}|^2 d\nu = \frac{\pi}{2\hbar} |H'_{\nu n}(0)|^2 \delta(E_\nu - E_n^{(0)} - \hbar\omega) d\nu. \quad (56.8)$$

In this case the wave functions of the continuous spectrum must be normalized to the δ -function in the ν -space. Formula (56.8) shows that under the action of a perturbation harmonically dependent on time the system may carry out transitions only into states with energy $E_\nu = E_n^{(0)} + \hbar\omega$.

The transition probability is defined by the square of the matrix element of the perturbation operator and depends, of course, on the choice of quantities characterizing the state of the continuous spectrum. The energy of the particle is often chosen as one of the parameters characterizing a state in the continuous spectrum. Then, integrating with respect to other parameters, we have

$$dW_{En} = \frac{\pi}{2\hbar} |H'_{En}(0)|^2 \rho(E) \delta(E - E_n^{(0)} - \hbar\omega) dE, \quad (56.9)$$

where $\rho(E) dE$ is the number of states with an energy in the interval from E to $E + dE$, and the following notation is introduced:

$$dE \int |H'_{\nu n}|^2 \frac{d\nu}{dE} = |H'_{En}|^2 \rho(E) dE.$$

Integrating over energy we find the total probability of the transition per unit time from a state with energy $E_n^{(0)}$ to a state of the continuous spectrum under the action of a harmonic perturbation:

$$W = \frac{\pi}{2\hbar} |H'_{En}(0)|^2 \rho(E), \quad (56.10)$$

* See L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

where $E = E_n^{(0)} + \hbar\omega$. We note that if, as distinct from (56.1), we denote the matrix element of the perturbation, introducing exponential functions, by

$$H'_{vn}(t) = H'_{vn}(0) (e^{i\omega t} + e^{-i\omega t}),$$

then the numerical coefficient in formulae (56.8)–(56.10) would evidently change by a factor of 4. For example, formula (56.8) would then be rewritten in the form

$$dW_{vn} = \frac{2\pi}{\hbar} |H'_{vn}(0)|^2 \delta(E_v - E_n^{(0)} - \hbar\omega) dv \quad (56.8')$$

and formulae (56.9) and (56.10) change analogously.

Another particularly important case is the transition caused by a time-independent perturbation. The expression for the transition probability can be obtained from formula (56.8) by setting the frequency in it to $\omega = 0$ and doubling the matrix element of the perturbation. This is associated with the fact that a term, which for $\omega = 0$ is the same as the term retained, was dropped in the transition from (56.2) to (56.3). For the transition probability we have

$$dW_{v_0} = \frac{2\pi}{\hbar} |H'_{v_0}|^2 \delta(E_v - E_{v_0}) dv. \quad (56.11)$$

A time-independent perturbation can give rise to transitions only to states with the same energy. In other words, it can cause transitions only between degenerate states. We have here denoted the initial state by the index v_0 , since transitions in the continuous spectrum are of the greatest interest in the case of the action of a constant perturbation. Of course, all the above reasoning associated with the transition to the δ -function is also valid in this case.

Integrating (56.11) over final state energies, we can write the transition probability in another form:

$$dW_{v_0} = \frac{2\pi}{\hbar} \int |H'_{v_0}|^2 \frac{dv}{dE} \delta(E - E_{v_0}) dE = \frac{2\pi}{\hbar} |H'_{v_0}|^2 \frac{dv}{dE}. \quad (56.12)$$

We write the total transition probability, analogously to (56.10), in the form

$$W = \frac{2\pi}{\hbar} |H'_{Ev_0}|^2 \rho(E), \quad (56.13)$$

where $E = E_{v_0}$. Let, for example, the final state be characterized by defining the momentum of the particle, so that

$$dv = dp_x dp_y dp_z = p^2 dp d\Omega = pm dE d\Omega,$$

where $E = \mathbf{p}^2/2m$ is the energy of the final state of the particle, and $d\Omega$ is an element of solid angle. Formula (56.12) is, in this case, rewritten in the form

$$dW_{\mathbf{p}v_0} = \frac{2\pi}{\hbar} |H'_{\mathbf{p}v_0}|^2 m p d\Omega, \quad \text{where } p = (2mE_{v_0})^{1/2}. \quad (56.14)$$

Here the wave functions of the final state must be normalized to the δ -function in momentum space.

In another method for the normalization of these functions, for example normalization in a "box" (see (26.16) and (26.17)), the interval of final states dv' will have the form

$$dv' = dn_x dn_y dn_z = \frac{dp_x dp_y dp_z V}{(2\pi\hbar)^3}. \quad (56.15)$$

Of course the expression for the transition probability (56.14) will not change in this case. Finally, we note that expressions (56.11)–(56.14) depend on the method of normalizing the wave function of the initial state, which also belongs to the continuous spectrum.

The matrix element of the perturbation operator very often turns out to be equal to zero. In this case the transition probability reduces to zero. This means that the corresponding transition is impossible in the first approximation of perturbation theory. In the next higher approximation the probability of the corresponding transition may turn out to be different from zero.

Let us find the probability of transition caused by a time-independent perturbation in the second order approximation for such a case.

Formula (55.11) gives

$$c_v^{(2)} = \frac{1}{i\hbar} \sum_k \int_0^t H'_{vk} c_k^{(1)}(t) \exp(i\omega_{vk}t) dt. \quad (56.16)$$

The sum (or over the continuous spectrum, the integral) involved here is taken over intermediate or, as they are often called, virtual states, so that the transition itself can be treated as a transition via intermediate states. It should be stressed that the transition of the system via intermediate states is not a real physical process, but serves only as a way of dealing with the formulae. Hence, for example, in transitions into virtual states the energy of the system does not need to be conserved. Substituting the expression for $c_k^{(1)}$ from (56.2) ($\omega = 0$) into (56.16) and integrating, we obtain

$$c_v^{(2)} = \frac{1}{\hbar^2} \sum_k H'_{vk} H'_{kv_0} \left[\frac{\exp(i\omega_{vv_0}t) - 1}{\omega_{kv_0} \omega_{vv_0}} - \frac{\exp(i\omega_{vk}t) - 1}{\omega_{kv_0} \omega_{vk}} \right]. \quad (56.17)$$

Since, by assumption, transitions are absent in the first approximation of perturbation theory, the matrix element of the perturbation operator $H'_{kv_0} = 0$ for transitions proceeding with energy conservation $\omega_{kv_0} = 0$. In correspondence with this those intermediate states for which $\omega_{kv_0} = 0$ give no contribution to the amplitude (56.17). For transitions proceeding with energy conservation* ($\omega_{vv_0} = 0$) the second term in the bracket of formula (56.17) is not large. Indeed, it might give an appreciable contribution only for $\omega_{vk} = 0$. But $\omega_{vv_0} = \omega_{vk} + \omega_{kv_0}$, and for $\omega_{vv_0} = 0$ and $\omega_{vk} = 0$ it turns out that also ω_{kv_0} reduces to zero. For such transitions $H'_{kv_0} = 0$ and, consequently, they can be disregarded. Proceeding from this, we can rewrite (56.17) as

$$c_v^{(2)} = -\frac{1}{\hbar} \Lambda_{vv_0} \frac{\exp(i\omega_{vv_0}t) - 1}{\omega_{vv_0}}, \quad (56.18)$$

where

$$\Lambda_{vv_0} = \sum_k \frac{H'_{vk} H'_{kv_0}}{E_{v_0} - E_k} \quad (56.19)$$

(integration is implied over the continuous spectrum).

We see that in the notation in the form of (56.18) the expression for the amplitude $c_v^{(2)}$ is the same as (56.2) ($\omega = 0$). Therefore the results obtained, in particular the formula for the transition probability (56.8), are conserved under the condition that the matrix element H'_{vv_0} be replaced by the matrix element Λ_{vv_0} .

§57. The adiabatic theory of perturbations

In certain cases the perturbation acting on a quantum system is associated with a slow, adiabatic change of the parameters on which the state of the system depends.

In the case of an adiabatic change of certain of the parameters which characterize a system it turns out that it is possible to develop a special approximate method of calculation called the adiabatic theory of perturbations. We shall encounter this method below in studying the properties of molecules and solid bodies. In such systems there are particles of two kinds:

* The possibility of transitions with non-conservation of energy is associated with the assumption of suddenness of switching on the perturbation (see (56.1)). For a more detailed discussion of this problem see, for example, L. Schiff, *Quantum mechanics* (McGraw-Hill, New York, 1949).

light electrons moving with large velocities, and heavy nuclei performing relatively slow movements. We shall call the electrons of the system the fast sub-system, and the heavy nuclei the slow sub-system. Roughly speaking, the characteristic time needed for a change of state of the fast sub-system is very small in comparison with the corresponding time for the slow sub-system. The essence of the adiabatic theory of perturbations amounts to the fact that the motion of the fast sub-system is considered in the first order approximation for given coordinates of the slow sub-system.

In other words, the motions of the fast and slow sub-systems are to a certain degree independent.

Let us consider the motion of a system consisting of electrons and nuclei. We write the Schrödinger equation in the form

$$\left[-\frac{\hbar^2}{2M} \sum_i \nabla_i^2 - \frac{\hbar^2}{2m} \sum_k \nabla_k^2 + U(\mathbf{r}_k, \mathbf{R}_i) \right] \psi(\mathbf{r}_k, \mathbf{R}_i) = E \psi(\mathbf{r}_k, \mathbf{R}_i). \quad (57.1)$$

Here m and M are the masses of the electrons and nuclei respectively. The sum over k is carried out with respect to the coordinates of the electrons, while the sum over i corresponds to the coordinates of the nuclei. U is the operator corresponding to the mutual interaction energy of the particles.

Further, we assume that it is possible to find the solution of the following Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m} \sum_k \nabla_k^2 + U(\mathbf{r}_k, \mathbf{R}_i) \right] \varphi_n(\mathbf{r}_k, \mathbf{R}_i) = E_n(\mathbf{R}_i) \varphi_n(\mathbf{r}_k, \mathbf{R}_i). \quad (57.2)$$

Equation (57.2) has the following physical meaning. The nuclei are assumed to be fixed at points \mathbf{R}_i . Finding the solution of eq. (57.2) comes down to the determination of the electron wave function φ_n and the energy levels of the electron system. As is seen from eq. (57.2), the energy levels $E_n(\mathbf{R}_i)$ of the electron sub-system depend on the coordinates of the nuclei (the heavy sub-system) as parameters.

Geometrically the electron energy $E_n(\mathbf{R}_i)$ forms a certain surface in space \mathbf{R}_i . This surface is called the electron term.

We write the solution of eq. (57.1) in the form of an expansion in a series in terms of the complete system of wave functions φ_n ,

$$\psi(\mathbf{r}_k, \mathbf{R}_i) = \sum_n \alpha_n(\mathbf{R}_i) \varphi_n(\mathbf{r}_k, \mathbf{R}_i). \quad (57.3)$$

We substitute (57.3) into (57.1), and then multiply eq. (57.1) by φ_m^* and integrate with respect to the coordinates of the electrons $dV = dV_1 dV_2 \dots$.

Taking into account the formula

$$\nabla_i^2 \alpha_n \varphi_n = \varphi_n \nabla_i^2 \alpha_n + \alpha_n \nabla_i^2 \varphi_n + 2 \nabla_i \varphi_n \cdot \nabla_i \alpha_n$$

we find the following equation:

$$\begin{aligned} & -\frac{\hbar^2}{2M} \sum_i \nabla_i^2 \alpha_m + E_m(\mathbf{R}_i) \alpha_m = \\ & = E \alpha_m + \sum_n \sum_i \left[\frac{\hbar^2}{2M} \alpha_n \int \varphi_m^* \nabla_i^2 \varphi_n dV + \right. \\ & \quad \left. + \frac{\hbar^2}{M} \int \varphi_m^* \nabla_i \varphi_n \cdot \nabla_i \alpha_n dV \right]. \end{aligned} \quad (57.4)$$

Here ∇_i is calculated with respect to the coordinates \mathbf{R}_i of the nuclei. We rewrite eq. (57.4) in the form

$$\left[-\frac{\hbar^2}{2M} \sum_i \nabla_i^2 + E_m(\mathbf{R}_i) \right] \alpha_m(\mathbf{R}_i) = E \alpha_m(\mathbf{R}_i) + \hat{C} \alpha_m, \quad (57.5)$$

where the operator \hat{C} is defined in the following way:

$$\hat{C} \alpha_m = \sum_i \sum_n \left(\frac{\hbar^2}{M} \nabla_i \alpha_n \int \varphi_m^* \nabla_i \varphi_n dV + \frac{\hbar^2}{2M} \alpha_n \int \varphi_m^* \nabla_i^2 \varphi_n dV \right). \quad (57.6)$$

The operator \hat{C} is called the non-adiabatic operator.

If one assumes that the operator \hat{C} is small and neglects it in eq. (57.5) then the equations for the functions φ_m and α_m assume the form

$$\left[-\frac{\hbar^2}{2m} \sum_k \nabla_k^2 + U(\mathbf{r}_k, \mathbf{R}_i) \right] \varphi_m = E_m(\mathbf{R}_i) \varphi_m, \quad (57.7)$$

$$\left[-\frac{\hbar^2}{2M} \sum_i \nabla_i^2 + E_m(\mathbf{R}_i) \right] \alpha_m = E \alpha_m. \quad (57.8)$$

Thus we obtain an important result in the zero order approximation with respect to the operator \hat{C} . Equation (57.7) represents a Schrödinger equation. The coordinates of the nuclei are involved in this equation as parameters. The function $\varphi_m(\mathbf{r}_k, \mathbf{R}_i)$ describes the motion of the electrons for motionless nuclei. Equation (57.8) contains only operators acting on the coordinates of the nuclei. It can be considered as the Schrödinger equation for the heavy sub-system, the nuclei. Then the energy $E_m(\mathbf{R}_i)$ of the electron sub-system plays the role of the potential energy of the nuclei.

The total wave function of the system in the zero order approximation

$\hat{C} = 0$ can be written in the form of a product of the wave functions α_m and φ_m , i.e. it has the same form as if the two sub-systems were quite independent:

$$\psi = \alpha_m(\mathbf{R}_i) \varphi_m(\mathbf{r}_k, \mathbf{R}_i).$$

In the approximation described it can be said that the electron sub-system follows the motion of the nuclei adiabatically in the sense that the electron sub-system remains in the same quantum state E_m when the position \mathbf{R}_i of the nuclei is changed. However, its energy level E_m changes in correspondence with the motion of the nuclei.

The condition of the smallness of the operator \hat{C} cannot be formulated in general form. In every actual problem this condition must be considered separately. Examples of such a consideration can be found in the books of Pauli, and Born and Huan Kun*.

§58. Perturbation theory in integral form

Perturbation theory can easily be developed within the framework of the Feynman formalism**. For this it is convenient to use as a basis the integral equation (29.5) for the Green's function $K(\mathbf{r}_2 t_2; \mathbf{r}_1 t_1)$ which we shall denote by $K(2, 1)$

$$K(2, 1) = K_0(2, 1) - \frac{i}{\hbar} \int_{-\infty}^{\infty} K_0(2, 3) \hat{H}'(3) K(3, 1) d^4 x_3. \quad (58.1)$$

Here we have denoted the Green's function of the unperturbed problem $\hat{H} = \hat{H}_0$, $\hat{H}' = 0$ by $K_0(2, 1)$.

Making use of the smallness of perturbation, we solve eq. (58.1) by a method of successive approximations. In the zeroth approximation, i.e. assuming $\hat{H}' = 0$, we have

$$K(2, 1) = K_0(2, 1). \quad (58.2)$$

* W. Pauli, *Die allgemeinen Prinzipien der Wellenmechanik (General principles of wave mechanics)*, Handbuch der Physik V/1, 1958; M. Born and Huang Kun, *Dynamical theory of crystal lattices* (University Press, Oxford, 1954).

** R. P. Feynman, Phys. Rev. 76 (1949) 740. See also S. Schweber, H. Bethe and F. Hofman, *Mesons and fields* (Row, Peterson and Co., Evanston, Ill. and White Plains, N.Y., 1956).

We shall obtain the next approximation if we substitute into the integral (58.1) the zeroth order approximation of the Green's function $K(3, 1)$, i.e.

$$K^{(1)}(2, 1) = -\frac{i}{\hbar} \int K_0(2, 3) \hat{H}'(3) K_0(3, 1) d^4x_3. \quad (58.3)$$

To obtain the correction to the Green's function in the second order approximation, we have to substitute into integral (58.1) the function $K(3, 1)$ with an accuracy to within terms of the first order of small quantities:

$$K^{(2)}(2, 1) = \left(-\frac{i}{\hbar}\right)^2 \int K_0(2, 3) \hat{H}'(3) K_0(3, 4) \hat{H}'(4) K_0(4, 1) d^4x_3 d^4x_4. \quad (58.4)$$

The correction to any order can be obtained in an analogous way. Finally, we have

$$\begin{aligned} K(2, 1) = & K_0(2, 1) - \frac{i}{\hbar} \int K_0(2, 3) \hat{H}'(3) K_0(3, 1) d^4x_3 + \\ & + \left(-\frac{i}{\hbar}\right)^2 \int K_0(2, 3) \hat{H}'(3) K_0(3, 4) \hat{H}'(4) K_0(4, 1) d^4x_3 d^4x_4 + \dots \end{aligned} \quad (58.5)$$

Formula (58.5) can be interpreted in the following way. The zero order term describes the motion of the unperturbed particle from point 1 to point 2. The next term describes the motion of the free particle from point 1 to point 3. At point 3 a perturbation acts. Thereupon the particle, again as a free particle, moves from point 3 to point 2. The integration means that we sum the contribution of all possible points 3. The term of second order smallness takes into account the action of the perturbation at two points, 3 and 4, and so on. By calculating the Green's function K from eq. (58.5) to a given approximation, we also know the wave function in this approximation. The convenience of the integral equation (58.1) lies in the fact that it makes it possible to obtain very simple a perturbation theory series. Examples of the use of the integral from the perturbation theory will be considered in ch. 14.

Spin and Identity of Particles

§59. The spin of elementary particles

Up to now we have assumed that the state of an individual microparticle is defined if its three space coordinates, or three momentum components, or in general three quantities forming a complete set are known. A number of experimental results indicated that many microparticles, for example electrons, protons, neutrons, have a specific intrinsic degree of freedom. This intrinsic degree of freedom is associated with an intrinsic angular momentum of the particle which does not depend on its orbital motion. This angular momentum of the particle is called the spin. The fact that the electron has a spin was established before the development of quantum mechanics. Attempts were made to interpret the spin as a manifestation of the rotation of a particle about its own axis (whence its name arose). However, this classical interpretation turned out to be untenable. All attempts to obtain the correct value of the ratio of the angular momentum to the magnetic moment for a system of a distributed rotating charge failed. As for the model of a rigid rotating sphere, (for which any value of this ratio can be obtained), such a model, as was explained in §13 of Part II, contradicts the general propositions of the theory of relativity. This contradiction was resolved in quantum mechanics. As we shall see below, this intrinsic degree of freedom and the spin associated with it have a specific quantum character. In the transition $\hbar \rightarrow 0$ to classical

mechanics the spin reduces to zero. Hence the spin has no classical analogue and does not allow interpretations of a classical character. The hypothesis of the existence of spin was initially put forward in connection with the interpretation of the spectra of alkali metals. Subsequently a number of facts were established enabling one to state unambiguously that this hypothesis is correct.

In the experiments of Stern and Gerlach the magnetic moment which is not associated with the orbital motion of electrons was observed directly. Namely, in these experiments it was established that when a beam of hydrogen atoms in an S-state was passed through a non-uniform magnetic field, then this beam split into two beams. However, in the S-state there is no orbital angular momentum, and consequently no orbital magnetic moment, so that the beam should pass through the magnetic field without undergoing any deflection.

The two-fold splitting is indicative of two possible orientations of the magnetic moment of the electron. The value of the spin magnetic moment can be determined from the magnitude of the splitting.

Direct experiments carried out by Einstein and de Haas made it possible to determine the ratio of the intrinsic angular momentum to the magnetic moment.

The spin of the electron (the intrinsic angular momentum) possesses the general properties of a quantum-mechanical moment which were discussed in §51. This was proved rigorously by the mathematical technique of group theory. In particular, the eigenvalue of the operator of the square of the spin moment $\hat{s}^2 = \hat{s}_x^2 + \hat{s}_y^2 + \hat{s}_z^2$ is expressed by the formula

$$s(s+1)\hbar^2, \quad (59.1)$$

where s denotes the corresponding quantum number, the intrinsic or spin quantum number of the particles. This quantum number is often called briefly the spin of the particle.

The number of possible spin projections onto an arbitrarily oriented z -axis is equal to $2s + 1$. The value of the intrinsic number s for each elementary particle must be determined experimentally. For the electron the existence and the value of the spin follows strictly from relativistic quantum mechanics, (Dirac's theory) to which ch. 13 is devoted.

The spins of the elementary particles which are most often encountered are the following: for the electron $s = \frac{1}{2}$, for the proton and neutron $s = \frac{1}{2}$, for the π -meson $s = 0$, for the muon $s = \frac{1}{2}$. This means that the possible values of the projections of the intrinsic angular momentum onto an axis arbitrarily

oriented in space, for example for the electron and other particles with spin $\frac{1}{2}$, are

$$s_z = \pm \frac{1}{2} \hbar. \quad (59.2)$$

From the experiment of Stern and Gerlach, the corresponding projections of the intrinsic magnetic moment of the electron are equal in absolute value to the Bohr magneton μ_0

$$\mu_z = \mp \frac{|e| \hbar}{2mc} = \mp \mu_0. \quad (59.3)$$

It is of great importance that the ratio of the intrinsic magnetic moment to the intrinsic angular momentum is equal to e/mc

$$\mu = \frac{e}{mc} s, \quad (59.4)$$

whereas for the orbital motion this ratio is smaller by a factor of 2 (see §63).

In §118 we shall show that this value of the intrinsic magnetic moment can be derived theoretically from Dirac's relativistic wave equation.

The spin properties of elementary particles play a very important role in the realm of microphenomena as well as in the behaviour of macroscopic bodies. This is associated with the fact that the spin determines directly the statistical properties of systems of quantum particles.

§60. Spin operators

Although, as we shall see below, the existence of spin for the electron and all the properties associated with it can be established theoretically from the propositions of relativistic quantum mechanics, a number of the properties of particles having spin can also be obtained without referring to relativistic theory, on the basis of general quantum-mechanical considerations and a relatively small number of experimental data. Since such a semi-empirical theory of particles with spin has a rather simple character but still makes it possible to obtain important results, we shall dwell on it below.

The wave function of a particle with spin depends not only on its three spatial coordinates but also on a fourth coordinate characterizing the intrinsic state of the particle. The value of the spin projection onto an arbitrarily oriented z -axis in space can be chosen as the fourth coordinate. Then the wave function can be written in the form

$$\psi = \psi(x, s_z, t). \quad (60.1)$$

As distinct from the spatial coordinates x , the 'spin coordinate' s_z takes on only a discrete sequence of values. The number of possible values of s_z is determined by the properties of the given elementary particle. As was mentioned above, the spin of most elementary particles is equal to one half. Since in this case the spin projection can take on only two values, the wave function (60.1) is conveniently written in the form of a column with two rows:

$$\psi = \begin{pmatrix} \psi(x, \frac{1}{2}\hbar, t) \\ \psi(x, -\frac{1}{2}\hbar, t) \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}. \quad (60.2)$$

We can then interpret $|\psi_1|^2 dV$ as the probability that the electron at the instant of time t will be found in the volume element dV and that it will have a spin component along the z -axis equal to $\frac{1}{2}\hbar$. Correspondingly, $|\psi_2|^2 dV$ is the probability that for the electron found in the volume element dV the spin component along the z -axis is equal to $-\frac{1}{2}\hbar$. The wave function $\psi(x, s_z, t)$ is assumed to be normalized in such a way that

$$\sum_{s_z} \int |\psi(x, s_z, t)|^2 dV = 1,$$

where the sum is taken over all possible values of the spin projection s_z . If the probability of one or another spin projection does not depend on the coordinates of the particle, then the wave function (60.2) can be rewritten in the form

$$\psi = \psi(x, t)\varphi, \quad (60.3)$$

where $\psi(x, t)$ is the ordinary (coordinate) wave function, $\varphi = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ is the spin wave function, and c_1 and c_2 are numbers. $|c_1|^2$ and $|c_2|^2$ give the probabilities that the spin projection s_z is equal respectively to $+\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$.

By virtue of the normalization condition for the wave function, we have

$$|c_1|^2 + |c_2|^2 = 1. \quad (60.4)$$

Having defined the concept of the spin wave function, we have to introduce the spin operators acting on it. In general the action of an operator on the spin function $\varphi = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ amounts to replacement of the components c_1 and c_2 by some linear combination of them

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \rightarrow \begin{pmatrix} a_{11}c_1 + a_{12}c_2 \\ a_{21}c_1 + a_{22}c_2 \end{pmatrix} \quad (60.5)$$

Corresponding with this the spin operator can be written in the form of a matrix

$$\hat{a} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}. \quad (60.6)$$

The action of such an operator on the wave function is given by formula (60.5), i.e.

$$\hat{a}\varphi = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} a_{11}c_1 + a_{12}c_2 \\ a_{21}c_1 + a_{22}c_2 \end{pmatrix}. \quad (60.7)$$

If the division of the wave function into a coordinate component and a spin component is not allowed, then formula (60.7) is rewritten in the form

$$\hat{a}\psi = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} a_{11}\psi_1 + a_{12}\psi_2 \\ a_{21}\psi_1 + a_{22}\psi_2 \end{pmatrix}. \quad (60.8)$$

The mean value of the operator \hat{a} taken in the state ψ is determined according to the general formula (44.8)

$$\bar{a}(x, t) = \psi_1^* a_{11} \psi_1 + \psi_1^* a_{12} \psi_2 + \psi_2^* a_{21} \psi_1 + \psi_2^* a_{22} \psi_2. \quad (60.9)$$

This equation can be rewritten in matrix form

$$\bar{a}(x, t) = \psi^\dagger \hat{a} \psi, \quad (60.10)$$

where ψ^\dagger is the matrix consisting of one row with the elements ψ_1^* and ψ_2^* :

$$\psi^\dagger = (\psi_1^* \psi_2^*). \quad (60.11)$$

Relation (60.10) determines the mean value of the quantity a at the instant of time t at the given point of space x . If this expression is averaged over all positions of the particle, then we obtain

$$\bar{a}(t) = \int \psi^\dagger \hat{a} \psi dV. \quad (60.12)$$

We now introduce the operators corresponding to the spin components \hat{s}_x , \hat{s}_y , \hat{s}_z . In §51 it was shown that the form of these operators and all the properties of the spin can be obtained if the commutation relations

$$\begin{aligned}
\hat{s}_x \hat{s}_y - \hat{s}_y \hat{s}_x &= i\hbar \hat{s}_z, \\
\hat{s}_y \hat{s}_z - \hat{s}_z \hat{s}_y &= i\hbar \hat{s}_x, \\
\hat{s}_z \hat{s}_x - \hat{s}_x \hat{s}_z &= i\hbar \hat{s}_y,
\end{aligned}
\tag{60.13}$$

are taken as the basis. The fact that the spin component operators must satisfy the same commutation relations as the operators of the components of the orbital angular momentum is, of course, not accidental. In §30 it was shown that the operator corresponding to the component of orbital angular momentum along any axis is associated with the operator corresponding to an infinitesimal rotation about this axis. The commutation relations (30.3) and (30.3') are a consequence of this fact, i.e. a consequence of the commutation relations between the infinitesimal rotation operators. In the next section we shall show that the spin component operators are also associated with rotation operators which act, however, not on the coordinate function but on the spin function. The commutation relations (60.13) are a consequence of the commutation relations between the infinitesimal rotation operators about the x-axis, y-axis and z-axis. The above considerations are rigorously substantiated by the methods of group theory*.

By analogy with (30.4), it follows from the relations (60.13) that

$$\begin{aligned}
\hat{s}_x \hat{s}^2 - \hat{s}^2 \hat{s}_x &= 0, \\
\hat{s}_y \hat{s}^2 - \hat{s}^2 \hat{s}_y &= 0, \\
\hat{s}_z \hat{s}^2 - \hat{s}^2 \hat{s}_z &= 0,
\end{aligned}
\tag{60.14}$$

Thus the square of the total spin and one of its projections onto an arbitrary axis can be measured simultaneously. The other two projections have no simultaneously sharp values.

For $s = \frac{1}{2}$ the matrices corresponding to the total spin, $\hat{s}^2 = \hat{s}_x^2 + \hat{s}_y^2 + \hat{s}_z^2$, and its projection on the z-axis have, in their own representation, the form

$$\hat{s}^2 = \frac{3}{4}\hbar^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \quad \hat{s}_z = \frac{1}{2}\hbar \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{1}{2}\hbar \sigma_z, \tag{60.15}$$

(the diagonal matrix elements are equal to the eigenvalues of the corresponding operators). According to the general formula (51.16), the matrices \hat{s}_x , \hat{s}_y in this representation are written as

* W. Pauli, *Die allgemeinen Prinzipien der Wellenmechanik (General principles of wave mechanics)*, Handbuch der Physik V/1, (Springer, Berlin, 1958).

$$\hat{s}_x = \frac{1}{2}\hbar \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{1}{2}\hbar \sigma_x; \quad \hat{s}_y = \frac{1}{2}\hbar \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \frac{1}{2}\hbar \sigma_y. \quad (60.16)$$

The matrices $\sigma_x, \sigma_y, \sigma_z$, which differ from the matrices $\hat{s}_x, \hat{s}_y, \hat{s}_z$ by a constant factor $\frac{1}{2}\hbar$, are called the Pauli matrices. They satisfy the following commutation relations:

$$\left. \begin{aligned} \sigma_x \sigma_y &= -\sigma_y \sigma_x = i\sigma_z, \\ \sigma_y \sigma_z &= -\sigma_z \sigma_y = i\sigma_x, \\ \sigma_z \sigma_x &= -\sigma_x \sigma_z = i\sigma_y, \end{aligned} \right\} \quad (60.17)$$

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1.$$

An arbitrary matrix of the second rank can be expressed in terms of the matrices $\sigma_x, \sigma_y, \sigma_z$ and the unit matrix.

As well as the similarity between the orbital and intrinsic angular momenta there is also a basic difference between them. The orbital angular momentum is characterized by the quantum number l which can take on any integer values irrespective of the nature of the particle, whereas the spin quantum number s takes on a limited number of values, for example $s = \frac{1}{2}$ for most elementary particles. Every kind of elementary particles has its own characteristic value of the spin. If the transition to classical mechanics is made by assuming $\hbar \rightarrow 0$, then, as was explained in §41, one must pass simultaneously to the limit of large quantum numbers. Hence, although according to formula (30.15) $l^2 = \hbar^2 l(l+1)$, it still does not follow that as $\hbar \rightarrow 0$ $k \rightarrow 0$, because simultaneously with $\hbar \rightarrow 0$ one has to assume that $l \rightarrow \infty$. In the case of the intrinsic angular momentum the situation is different. Since s takes on only a limited number of values, the transition to classical mechanics always leads to the spin value $s = 0$. We see that in classical mechanics there is no quantity which could serve as the classical analogue of the spin. The spin is a purely quantum concept characterizing the specific properties of microparticles.

§61. The eigenfunctions of the operators of the components of the spin of a particle. The rotation matrix

Let us find the eigenfunctions and eigenvalues of the operators $\hat{s}_x, \hat{s}_y, \hat{s}_z$. The equation for the eigenfunctions $\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ and the eigenvalues s_x of the operator \hat{s}_x has the form

$$\hat{s}_x \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = s_x \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Taking into account (60.16) and carrying out the multiplication, we obtain

$$\begin{pmatrix} \frac{1}{2}\hbar c_2 \\ \frac{1}{2}\hbar c_1 \end{pmatrix} = \begin{pmatrix} s_x c_1 \\ s_x c_2 \end{pmatrix}.$$

We evaluate this equality:

$$\frac{1}{2}\hbar c_2 = s_x c_1, \quad \frac{1}{2}\hbar c_1 = s_x c_2. \quad (61.1)$$

Hence, upon multiplying, we find $s_x = \pm \frac{1}{2}\hbar$.

The eigenvalues of the spin component operator, as was to be expected, turn out to be equal to $\pm \frac{1}{2}\hbar$. We determine the form of the eigenfunctions corresponding to these eigenvalues.

For $s_x = +\frac{1}{2}\hbar$ we have from (61.1) $c_1 = c_2$. Taking into account the normalization condition (60.4), we finally get

$$\varphi_{s_x=+\hbar/2} = \frac{1}{\sqrt{2}} e^{i\alpha_1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad (61.2)$$

where α_1 is an arbitrary phase.

Correspondingly, for $s_x = -\frac{1}{2}\hbar$, $c_1 = -c_2$, and the spin wave function is written in the form

$$\varphi_{s_x=-\hbar/2} = \frac{1}{\sqrt{2}} e^{i\alpha_2} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \quad (61.3)$$

Naturally, the eigenvalues of the operators \hat{s}_y and \hat{s}_z are also equal to $\pm \frac{1}{2}\hbar$. We also find their eigenfunctions in an analogous way

$$\varphi_{s_y=\hbar/2} = \frac{1}{\sqrt{2}} e^{i\alpha_3} \begin{pmatrix} 1 \\ i \end{pmatrix}; \quad \varphi_{s_y=-\hbar/2} = \frac{1}{\sqrt{2}} e^{i\alpha_4} \begin{pmatrix} 1 \\ -i \end{pmatrix}, \quad (61.4)$$

$$\varphi_{s_z=\hbar/2} = \varphi_{\frac{1}{2}} = e^{i\alpha_5} \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \quad \varphi_{s_z=-\hbar/2} \equiv \varphi_{-\frac{1}{2}} = e^{i\alpha_6} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (61.5)$$

The arbitrary phase factors α_i can, in particular, be set equal to zero.

We now consider a certain rotation of the system of coordinates $x, y, z \rightarrow x', y', z'$. Then the spin wave functions also change, $\varphi \rightarrow \varphi'$. Indeed, the transition from the coordinate system x, y, z to the system x', y', z' means a corre-

sponding transition from one representation to another. Such a transition, as we know (see §46), is carried out by means of a certain unitary matrix \hat{T} , so that $\varphi' = \hat{T}\varphi$. In the given case it is natural to call the unitary matrix \hat{T} the rotation matrix. Let us define this matrix. We first consider a rotation about the z-axis through an angle η . According to (46.15), the operators of the spin components in the new representation have the form

$$\begin{aligned}\hat{s}'_x &= \hat{T}_z \hat{s}_x \hat{T}_z^{-1}, \\ \hat{s}'_y &= \hat{T}_z \hat{s}_y \hat{T}_z^{-1}, \\ \hat{s}'_z &= \hat{T}_z \hat{s}_z \hat{T}_z^{-1}.\end{aligned}\tag{61.6}$$

Here \hat{s}'_x , \hat{s}'_y , \hat{s}'_z are the operators of the spin projections onto the old coordinate axes x , y , z but taken in the new representation connected with the system x' , y' , z' . Since we consider a rotation about the z-axis, the z-axis and the z' -axis coincide and, consequently, the operators of the spin projections onto these axes are the same in the two representations (expressed by the diagonal matrix (60.15)). However, the conditions for the operator of the spin projection onto the z-axis to be chosen in the form of the diagonal matrix (60.15) are still insufficient for the definition of the operators of spin projections along other directions. The operator of the spin projection along an arbitrary direction will be known if we also give the operators \hat{s}_x and \hat{s}_y in the form of matrices, i.e. if we give the system of x, y, z coordinates with which we connect the representation. We can, in particular, choose the representation connected with the system x' , y' , z' . In this representation (which we denote by primes) the operators of spin projections onto the x' -axis, y' -axis and z' -axis $\hat{s}'_{x'}$, $\hat{s}'_{y'}$, $\hat{s}'_{z'}$ have the form (60.16), (60.15).

Because the operators \hat{s}_x , \hat{s}_y , \hat{s}_z correspond to the projections of the spin moment, when the coordinate system rotates they must transform as the projections of an angular momentum, i.e. as the components of an axial vector. Since we consider the rotation through an angle η about the z-axis, the operators of spin projections onto primed and non-primed coordinate axes are connected with each other by the relations

$$\begin{aligned}\hat{s}_x &= \hat{s}_{x'} \cos \eta - \hat{s}_{y'} \sin \eta, \\ \hat{s}_y &= \hat{s}_{x'} \sin \eta + \hat{s}_{y'} \cos \eta, \\ \hat{s}_z &= \hat{s}_{z'}.\end{aligned}\tag{61.7}$$

The equalities (61.7) hold in any representation. In particular, in the representation connected with the rotated system of coordinates we have

$$\begin{aligned}
\hat{s}'_x &= \hat{s}'_x \cos \eta - \hat{s}'_y \sin \eta = \hat{s}_x \cos \eta - \hat{s}_y \sin \eta, \\
\hat{s}'_y &= \hat{s}'_x \sin \eta + \hat{s}'_y \cos \eta = \hat{s}_x \sin \eta + \hat{s}_y \cos \eta, \\
\hat{s}'_z &= \hat{s}'_z = \hat{s}_z.
\end{aligned} \tag{61.8}$$

Indeed, the two systems of coordinates, the primed and the non-primed, are completely equivalent. As we have already noted, the operators of the spin projections onto the axes of the primed coordinate system, taken in the representation which is just connected with this rotated system of coordinates, have the ordinary form (60.15) and (60.16), i.e. are the same as the operators of the spin projections onto the axes of the non-primed system of coordinates, taken in the representation connected with the non-primed system of coordinates:

$$\hat{s}'_x = \hat{s}_x; \quad \hat{s}'_y = \hat{s}_y; \quad \hat{s}'_z = \hat{s}_z.$$

Considering the equalities (61.6) and (61.8) together, we find the matrix \hat{T}_z . First of all, it follows from (61.6) and (61.8) that $\hat{s}_z = \hat{T}_z \hat{s}_z \hat{T}_z^{-1}$, i.e. the matrix \hat{T}_z commutes with \hat{s}_z . Since the matrix \hat{s}_z is diagonal, the matrix \hat{T}_z is also diagonal (see §47). Consequently, the matrix \hat{T}_z has the form

$$\hat{T}_z = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}.$$

The condition for the unitarity of the matrix \hat{T}_z , $\hat{T}_z \hat{T}_z^\dagger = \hat{T}_z^\dagger \hat{T}_z = 1$, leads to the equality

$$\begin{pmatrix} |a_1|^2 & 0 \\ 0 & |a_2|^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

or $|a_1|^2 = 1$ and $|a_2|^2 = 1$. Consequently, the matrix \hat{T}_z has the form

$$\hat{T}_z = \begin{pmatrix} e^{i\alpha_1} & 0 \\ 0 & e^{i\alpha_2} \end{pmatrix}, \tag{61.9}$$

where α_1 and α_2 are real. We rewrite eq. (61.6) in the form

$$\hat{s}'_x \hat{T}_z = \hat{T}_z \hat{s}_x,$$

$$\hat{s}'_y \hat{T}_z = \hat{T}_z \hat{s}_y.$$

Substituting the values of \hat{s}'_x and \hat{s}'_y from (61.8) into these expressions and equating the corresponding matrix elements, we find that the equalities are satisfied under the condition $\alpha_1 - \alpha_2 = \eta$.

Thus for the two phases α_1 and α_2 we have obtained one condition connecting them. This fact is not accidental. The point is that the matrix \hat{T}_z can contain an arbitrary phase factor which will not affect the results because the wave functions themselves are defined to within an arbitrary phase factor. In accordance with this we write the matrix \hat{T}_z in the form

$$\hat{T}_z(\eta) = \begin{pmatrix} e^{\frac{1}{2}i\eta} & 0 \\ 0 & -e^{-\frac{1}{2}i\eta} \end{pmatrix}. \quad (61.10)$$

We can express the matrix \hat{T}_z also in terms of the matrix σ_z

$$\hat{T}_z(\eta) = \cos \frac{1}{2}\eta + i\sigma_z \sin \frac{1}{2}\eta, \quad (61.11)$$

or in a somewhat different form

$$\hat{T}_z = \exp\left(\frac{1}{2}i\eta\sigma_z\right). \quad (61.12)$$

The above expressions should be understood as

$$\hat{T}_z = 1 + \left(\frac{i}{2}\eta\sigma_z\right) + \frac{1}{2}\left(\frac{i}{2}\eta\sigma_z\right)^2 + \dots + \frac{1}{n!}\left(\frac{i}{2}\eta\sigma_z\right)^n + \dots$$

Since $\sigma_z^2 = 1$, $\sigma_z^3 = \sigma_z$ and so on, the series is easily evaluated, and we again arrive at formula (61.11). If the angle of rotation is small, then the rotation matrix (61.12) has the form

$$\hat{T}_z = 1 + i\frac{\eta}{\hbar}\hat{s}_z, \quad (61.13)$$

where we have introduced the matrix $\hat{s}_z = \frac{1}{2}\hbar\sigma_z$ instead of σ_z . We have obtained an expression which has the same structure as the rotation operator obtained in §30 which acts on a function depending on spatial coordinates.

Of course, a relation of the type (61.11) is also valid for a rotation about any other axis, since all directions are equivalent. Thus, for example, we write the matrix \hat{T}_x of a rotation about the x-axis through a certain angle ϑ :

$$\hat{T}_x(\vartheta) = \cos \frac{1}{2}\vartheta + i\sigma_x \sin \frac{1}{2}\vartheta. \quad (61.13')$$

The arbitrary rotation of one coordinate system with respect to another can be characterized by three Euler angles ϑ , φ , ψ (fig. V.18). The angle ψ is the angle between the axis Ox and the straight line ON forming the intersection of the planes xOy and x'Oy', ϑ is the angle between the axes Oz and Oz' and, finally, φ is the angle between ON and the axis Ox'. The total rotation can be resolved into three consecutive rotations: I about the z-axis through the angle ψ ; II about the new position of the axis x(ON) through the angle ϑ , and III

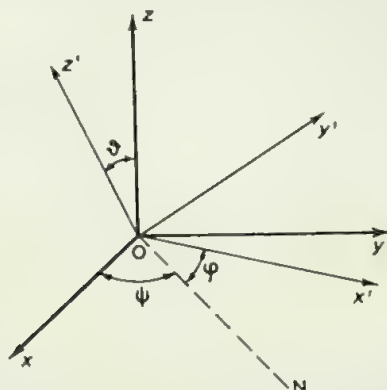


Fig. V.18

about the new position of the axis Oz through the angle φ . The rotation matrix \hat{T} will be equal to the product of three matrices $\hat{T} = \hat{T}_z(\varphi)\hat{T}_x(\theta)\hat{T}_z(\psi)$. Making use of (61.10) and (61.13') and multiplying the matrices, we find

$$\hat{T} = \begin{pmatrix} e^{\frac{1}{2}i(\psi+\varphi)} \cos \frac{1}{2}\theta & ie^{-\frac{1}{2}i(\psi-\varphi)} \sin \frac{1}{2}\theta \\ ie^{\frac{1}{2}i(\psi-\varphi)} \sin \frac{1}{2}\theta & e^{-\frac{1}{2}i(\psi+\varphi)} \cos \frac{1}{2}\theta \end{pmatrix}. \quad (61.14)$$

This matrix was first obtained by Pauli. The two-component wave function $\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ which transforms, when the system of coordinates rotates, according to the law

$$\begin{pmatrix} c'_1 \\ c'_2 \end{pmatrix} = \hat{T} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

is called a spinor.

The spinor components which we have introduced are usually denoted as

$$\varphi = \begin{pmatrix} \varphi^1 \\ \varphi^2 \end{pmatrix}.$$

We note that for any given spinor $\varphi = \begin{pmatrix} \varphi^1 \\ \varphi^2 \end{pmatrix}$ one can always find a matrix $\hat{T}(\vartheta, \varphi, \psi)$ such that $\varphi^{1'} = 1, \varphi^{2'} = 0$, i.e. it is always possible to determine the direction, characterized by the angles $(\varphi, \vartheta, \psi)$, along which the spin of the particle is oriented.

We rewrite the matrix \hat{T} in the form

$$\hat{T} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix},$$

so that

$$\begin{aligned} \varphi^{1'} &= \alpha\varphi^1 + \beta\varphi^2, \\ \varphi^{2'} &= \gamma\varphi^1 + \delta\varphi^2, \end{aligned} \quad (61.15)$$

where, as follows from (61.14),

$$\alpha = \delta^*, \quad \beta = -\gamma^*, \quad \alpha\delta - \beta\gamma = 1. \quad (61.16)$$

The transformation (61.15) is usually called a bilinear transformation.

The bilinear transformation leaves certain bilinear forms invariant. Indeed, making use of (61.15) and (61.16) we easily obtain for two arbitrary spinors φ and η

$$\begin{aligned} \varphi^{1'}\eta^{2'} - \varphi^{2'}\eta^{1'} &= (\alpha\delta - \beta\gamma)(\varphi^1\eta^2 - \varphi^2\eta^1) = \varphi^1\eta^2 - \varphi^2\eta^1, \\ \varphi^{1*}\varphi^1 + \varphi^{2*}\varphi^2 &= \text{const}. \end{aligned} \quad (61.17)$$

The above relation expresses the conservation of normalization under the rotation of a coordinate system.

By means of the spinor components η^1, η^2 and ξ^1, ξ^2 one can construct quantities which transform under a rotation of the coordinate system as the components of a vector, i.e. according to the law

$$a'_i = \sum_{k=1}^3 \alpha_{ik} a_k,$$

where α_{ik} are the cosines of the angles between the old and new coordinate axes. By a direct check, making use of (61.14) and (61.15), one can verify that the vector components are determined by the relations

$$\begin{aligned} a_z &= (\eta^1\xi^2 + \eta^2\xi^1), \\ a_x &= (\eta^2\xi^2 - \eta^1\xi^1), \\ a_y &= -i(\eta^1\xi^1 + \eta^2\xi^2). \end{aligned} \quad (61.18)$$

Correspondingly for the square of the vector we have

$$a^2 = a_x^2 + a_y^2 + a_z^2 = (\eta^1 \xi^2 - \eta^2 \xi^1)^2 \quad (61.19)$$

We have, as was to be expected, obtained a scalar quantity.

The components of a tensor of arbitrary rank $B_{ikl\dots}$ can be defined as the product $a_i b_k c_l \dots$ of the corresponding components of vectors. By means of formulae (61.18) we can identify the components $B_{ikl\dots}$ with the products of the components of spinors.

Finally, we shall define the law of transformation of a spinor under the inversion of the coordinate system, i.e. under the transformation $x \rightarrow -x$, $y \rightarrow -y$, $z \rightarrow -z$.

We denote the inversion operator by \hat{I} , so that

$$\varphi' = \hat{I}\varphi. \quad (61.20)$$

The transformation (61.20) can now be considered as a transition to another representation. The corresponding transformation of the operators \hat{s}_x , \hat{s}_y , \hat{s}_z is analogous to (61.6). On the other hand, we can consider the operators \hat{s}_x , \hat{s}_y , \hat{s}_z as the components of an axial vector (as the components of an orbital angular momentum). Consequently, these operators must not change sign under reflection. Based on this, we obtain, in analogy with formula (61.6)

$$\begin{aligned} \hat{s}_x &= \hat{I}\hat{s}_x\hat{I}^{-1}, \\ \hat{s}_y &= \hat{I}\hat{s}_y\hat{I}^{-1}, \\ \hat{s}_z &= \hat{I}\hat{s}_z\hat{I}^{-1}. \end{aligned} \quad (61.21)$$

When the matrix \hat{I} is applied twice we come back to the initial state and, consequently, we must obtain the spinor φ . Furthermore, we can obtain the spinor $-\varphi$ if we consider a double reflection as a rotation through the angle 2π . As is seen from formulae (61.10) and (61.13'), under such a rotation the spinor changes sign. Correspondingly we have

$$\hat{I}^2 = \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (61.22)$$

We see, consequently, that the matrix \hat{I} must commute with the matrices \hat{s}_x , \hat{s}_y , \hat{s}_z , and that its square must give the unit matrix multiplied by ± 1 . These requirements will be fulfilled under the condition that

$$\hat{I} = \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad \hat{I} = \pm i \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (61.23)$$

Consequently, under reflection the spinor can transform in the following way:

$$\varphi' = \pm \varphi \quad (61.24)$$

or

$$\varphi' = \pm i \varphi. \quad (61.25)$$

If the law of transformation is determined by the upper sign in formulae (61.24) and (61.25), then φ is sometimes called a polar spinor; if the law of transformation is determined by the lower sign, then φ is called a pseudo-spinor.

§62. The total angular momentum

The total angular momentum of a particle is made up of the orbital angular momentum and the spin angular momentum. According to the rules for the addition of vector operators we have for the total angular momentum operator $\hat{\mathbf{j}}$

$$\hat{\mathbf{j}} = \hat{\mathbf{l}} + \hat{\mathbf{s}}. \quad (62.1)$$

The orbital and intrinsic angular momentum operators act on different variables. The first acts on space variables, while the second acts only on spin variables. Hence the two operators commute with each other. It follows directly from this that the components of the total angular momentum operator satisfy the same commutation rules as the components of the orbital and intrinsic angular momenta. These commutation relations also follow from the connected between the total angular momentum operator and the rotation operator (see below)

$$\begin{aligned} \hat{j}_x \hat{j}_y - \hat{j}_y \hat{j}_x &= i\hbar \hat{j}_z, \\ \hat{j}_y \hat{j}_z - \hat{j}_z \hat{j}_y &= i\hbar \hat{j}_x, \\ \hat{j}_z \hat{j}_x - \hat{j}_x \hat{j}_z &= i\hbar \hat{j}_y, \end{aligned} \quad (62.2)$$

and also

$$\begin{aligned} \hat{j}_x \hat{j}^2 - \hat{j}^2 \hat{j}_x &= 0, \\ \hat{j}_y \hat{j}^2 - \hat{j}^2 \hat{j}_y &= 0, \\ \hat{j}_z \hat{j}^2 - \hat{j}^2 \hat{j}_z &= 0. \end{aligned} \quad (62.3)$$

It follows from the relations (62.2) (see §51) that the eigenvalues of the operator $\hat{\mathbf{j}}^2$ have the form

$$\mathbf{j}^2 = \hbar^2 j(j+1). \quad (62.4)$$

The value of the quantum number j determines the value of the total angular momentum. According to the rules of addition of angular momenta in quantum mechanics (52.1) it follows that for a given l and s the number j runs over the sequence of values

$$j = |l-s|, \quad |l-s| + 1, \dots, l+s-1, \quad l+s.$$

The quantum number j takes on integer values if the spin has integer values, and half-integer values if the spin is half-integer.

We shall show that for a particle moving in free space or in a centrally symmetric field it is the total angular momentum which is the constant of the motion. For proof we introduce the rotation operator \hat{R} taking into account the change in the spin coordinates as well the spatial coordinates of the wave function. We consider a rotation of the coordinate system about the z -axis through a small angle $\delta\varphi$. Proceeding from the results obtained in §30 and §61, it is easy to determine the change in the total wave function under such a rotation

$$\psi' = \hat{R}_z \psi = \hat{T}_z \hat{W}_z \psi = \left[1 + \frac{i}{\hbar} \delta\varphi (\hat{s}_z + \hat{l}_z) \right] \psi = \left(1 + \frac{i}{\hbar} \delta\varphi \hat{j}_z \right) \psi.$$

Of course, an analogous relation also holds for a rotation about any other axis. Consequently we see that it is simply the total angular momentum operator which is connected with the rotation operator. But the operation of rotation, by virtue of the isotropy of space, must not change the Hamiltonian of a closed system (or a system in a centrally symmetric field). Mathematically this is shown by the fact that the operator \hat{R}_z and, consequently, the operator \hat{j}_z will commute with the Hamiltonian \hat{H} of the particle. Thus the total angular momentum conservation law is a consequence of the isotropy of space. For the intrinsic and orbital angular momenta separately the conservation laws hold only approximately, to the extent that we neglect the spin-orbit interaction.

If we have a system of non-interacting particles, then the total angular momentum of the entire system $\hat{\mathbf{J}}$ is made up of the angular momenta $\hat{\mathbf{j}}_k$ of the individual particles according to the rules of addition of angular momenta in quantum mechanics

$$\hat{\mathbf{J}} = \sum \hat{\mathbf{j}}_k. \quad (62.5)$$

One can also introduce the total orbital angular momentum operator $\hat{\mathbf{L}}$

$$\hat{\mathbf{L}} = \sum_k \hat{\mathbf{l}}_k, \quad (62.6)$$

and the total spin operator of the system $\hat{\mathbf{S}}$

$$\hat{\mathbf{S}} = \sum_k \hat{\mathbf{s}}_k. \quad (62.7)$$

Since $\hat{\mathbf{j}}_k = \hat{\mathbf{l}}_k + \hat{\mathbf{s}}_k$, then, obviously, we have

$$\hat{\mathbf{J}} = \hat{\mathbf{L}} + \hat{\mathbf{S}}. \quad (62.8)$$

The operators referring to different particles commute with each other because they act on different variables. Hence for the operators of the components of the total angular momenta $\hat{\mathbf{J}}$, $\hat{\mathbf{L}}$, $\hat{\mathbf{S}}$ there are the same commutation relations as for the operators referring to individual particles. For example, for the operators \hat{J}_x , \hat{J}_y of the components of the total angular momentum we have

$$\begin{aligned} \hat{J}_x \hat{J}_y - \hat{J}_y \hat{J}_x &= \sum_{k,i} (\hat{J}_{kx} \hat{J}_{iy} - \hat{J}_{iy} \hat{J}_{kx}) = \\ &= \sum_k (\hat{J}_{kx} \hat{J}_{ky} - \hat{J}_{ky} \hat{J}_{kx}) = i\hbar \sum_k \hat{J}_{kz} = i\hbar \hat{J}_z. \end{aligned}$$

An analogous result is also obtained for other projections of the operator $\hat{\mathbf{J}}$, as well as for the operators $\hat{\mathbf{L}}$ and $\hat{\mathbf{S}}$.

For given eigenvalues of the operators referring to individual particles the eigenvalues of the operators $\hat{\mathbf{J}}^2$, $\hat{\mathbf{L}}^2$, $\hat{\mathbf{S}}^2$ and of the operators of the components of the angular momenta are determined according to the rules of addition of angular momenta in quantum mechanics. The total angular momentum \mathbf{J} is conserved quantity for a closed system of particles.

§63. The Pauli equation. The probability current density vector

In ch. 13, devoted to relativistic quantum mechanics, we shall establish the exact relativistic equation of quantum mechanics and show that the Schrödinger equation is obtained from it in the limit when $v/c \rightarrow 0$.

In this case it turns out that if terms of higher orders of small quantities are taken into account, then additional terms will arise in the Hamiltonian, describing a number of important properties of quantum systems.

In particular, the existence of spin, as well as the fact that the electron has a magnetic moment, follows from the relativistic wave equation by expanding in powers of v/c and preserving terms of the first order of small quantities.

Deferring the proof of this statement to §118, we introduce the intrinsic magnetic moment operator in correspondence with (59.4) by the relation

$$\hat{\boldsymbol{\mu}} = \alpha \hat{\mathbf{S}} = \frac{e}{mc} \hat{\mathbf{S}}. \quad (63.1)$$

Then the Hamiltonian for an electron in an electromagnetic field assumes the form

$$\begin{aligned} \hat{H} &= \frac{1}{2m} \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right)^2 + U(\mathbf{r}) - (\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{H}}) = \\ &= \frac{1}{2m} \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right)^2 + U(\mathbf{r}) - \frac{e}{mc} (\hat{\mathbf{S}} \cdot \boldsymbol{\mathcal{H}}), \end{aligned} \quad (63.2)$$

where $\boldsymbol{\mathcal{H}}$ is the magnetic field strength. Since the Hamiltonian depends on the spin, the wave function of the electron also depends on the spin variable, i.e. $\psi = \psi(x, y, z, t, s_z)$. The equation for the wave function $\psi(x, y, z, t, s_z)$ in a magnetic field, first introduced by Pauli and called the Pauli equation, is of the form

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{1}{2m} \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right)^2 \psi - \frac{e}{mc} (\hat{\mathbf{S}} \cdot \boldsymbol{\mathcal{H}}) \psi + U\psi. \quad (63.3)$$

Let us find the probability current density vector. For this we write the equation for the function ψ^\dagger

$$-i\hbar \frac{\partial \psi^\dagger}{\partial t} = \frac{1}{2m} \left(\hat{\mathbf{p}} + \frac{e}{c} \mathbf{A} \right)^2 \psi^\dagger - \frac{e}{mc} [(\hat{\mathbf{S}} \cdot \boldsymbol{\mathcal{H}}) \psi]^\dagger + U\psi^\dagger. \quad (63.3')$$

We multiply (63.3) from the left by ψ^\dagger and (63.3') from the right by ψ and subtract one from the other.

Taking into account that $\hat{\mathbf{p}} \cdot \mathbf{A} - \mathbf{A} \cdot \hat{\mathbf{p}} = (\hbar/i) \nabla \mathbf{A}$, we obtain after simple calculation

$$\begin{aligned} \frac{\partial(\psi^\dagger \psi)}{\partial t} &= -\frac{i\hbar}{2m} \nabla [(\nabla \psi^\dagger) \psi - \psi^\dagger \nabla \psi] + \\ &+ \frac{e}{mc} \nabla (\mathbf{A} \psi^\dagger \psi) + \frac{ie}{m\hbar} [\psi^\dagger (\hat{\mathbf{S}} \cdot \boldsymbol{\mathcal{H}}) \psi - ((\hat{\mathbf{S}} \cdot \boldsymbol{\mathcal{H}}) \psi)^\dagger \psi]. \end{aligned} \quad (63.4)$$

It follows from rule (45.20) that

$$((\hat{\mathbf{s}} \cdot \mathcal{H})\psi)^\dagger = \psi^\dagger (\hat{\mathbf{s}}^\dagger \cdot \mathcal{H}) \quad (63.5)$$

and, furthermore, $\hat{\mathbf{s}}^\dagger = \hat{\mathbf{s}}$ by virtue of the hermitian property of the spin operator. Thus the term in square brackets reduces to zero and we obtain an expression for the probability current density vector. (It should be borne in mind that it contains two-component wave functions). Multiplying it by the charge e , we obtain the electric current density vector

$$\mathbf{j} = \frac{i\hbar e}{2m} ((\nabla\psi^\dagger)\psi - \psi^\dagger \nabla\psi) - \frac{e^2}{mc} \mathbf{A} \psi^\dagger \psi. \quad (63.6)$$

The expression (63.4) determines the probability current density vector \mathbf{j} with an accuracy to within $\nabla \times \mathbf{B}$, where \mathbf{B} is an arbitrary vector. It can be shown that $\mathbf{B} = (e/m)\psi^\dagger \hat{\mathbf{s}}\psi$, so that the total expression for the electric current density will have the form*

$$\mathbf{j}_n = -\frac{i\hbar e}{2m} (\psi^\dagger \nabla\psi - (\nabla\psi^\dagger)\psi) - \frac{e^2}{mc} \mathbf{A}(\psi^\dagger \psi) + \frac{e}{m} \nabla \times (\psi^\dagger \hat{\mathbf{s}}\psi). \quad (63.7)$$

Let us consider the case where the particle moves in a constant magnetic field \mathcal{H} and the electric field is absent. Then the vector potential \mathbf{A} can be chosen in the form

$$\mathbf{A} = \frac{1}{2}(\mathcal{H} \times \mathbf{r}). \quad (63.8)$$

We transform the Pauli eq. (63.3), taking into account that the following relation holds for the vector potential (63.8)

$$\hat{\mathbf{p}} \cdot \mathbf{A} = \mathbf{A} \cdot \hat{\mathbf{p}}.$$

Furthermore we assume that the magnetic field \mathcal{H} is relatively weak, and in correspondence with this we drop terms in the Pauli equation which are quadratic in \mathcal{H} . We then have

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{\hat{\mathbf{p}}^2}{2m} \psi - \frac{e}{2mc} (\mathcal{H} \times \mathbf{r}) \cdot \hat{\mathbf{p}} \psi - \frac{e}{mc} (\hat{\mathbf{s}} \cdot \mathcal{H}) \psi.$$

Since

$$(\mathcal{H} \times \mathbf{r}) \cdot \hat{\mathbf{p}} = \mathcal{H} \cdot (\mathbf{r} \times \hat{\mathbf{p}}) = \mathcal{H} \cdot \hat{\mathbf{l}},$$

where $\hat{\mathbf{l}}$ is the orbital angular momentum, then

* See L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{\hat{\mathbf{p}}^2}{2m} \psi - (\hat{\boldsymbol{\mu}}_l \cdot \boldsymbol{\mathcal{H}}) \psi - (\hat{\boldsymbol{\mu}}_s \cdot \boldsymbol{\mathcal{H}}) \psi. \quad (63.9)$$

It is natural to call the operator $\hat{\boldsymbol{\mu}}_l$,

$$\hat{\boldsymbol{\mu}}_l = \frac{e}{2mc} \hat{\mathbf{L}},$$

the orbital magnetic moment operator (as distinct from $\hat{\boldsymbol{\mu}}_s$, the spin magnetic moment operator).

We see that the ratio of the orbital magnetic moment $\hat{\boldsymbol{\mu}}_l$ to the orbital angular momentum $\hat{\mathbf{L}}$ is, as in classical physics, equal to $e/2mc$. For spin moments this ratio is larger by a factor of two.

Eqs. (63.3) and (63.9) are naturally generalized to the case of a system of particles. Thus for a system of particles (of charge e and mass m) placed in a weak magnetic field $\boldsymbol{\mathcal{H}}$ the Pauli equation (63.9) has the form

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{1}{2m} \sum_k \hat{\mathbf{p}}_k^2 \psi - \frac{e}{2mc} (\hat{\mathbf{L}} \cdot \boldsymbol{\mathcal{H}}) \psi - \frac{e}{mc} (\hat{\mathbf{S}} \cdot \boldsymbol{\mathcal{H}}) \psi + U_{\text{int}} \psi, \quad (63.10)$$

where $\hat{\mathbf{L}} = \sum_k \hat{\mathbf{L}}_k$ is the operator of the total orbital angular momentum of the system, $\hat{\mathbf{S}} = \sum_k \hat{\mathbf{S}}_k$ is the operator of the total spin angular momentum of the system (the summation is carried out over all particles of the system), and U_{int} is the potential energy of interaction of the particles with each other.

The term which is quadratic in the magnetic field is also easily taken into account. In this case the Pauli equation has the form

$$i\hbar \frac{\partial \psi}{\partial t} = \left\{ \frac{1}{2m} \sum_k \hat{\mathbf{p}}_k^2 - \frac{e}{2mc} (\hat{\mathbf{L}} + 2\hat{\mathbf{S}}) \cdot \boldsymbol{\mathcal{H}} + \frac{e^2}{8mc^2} \sum_k [\boldsymbol{\mathcal{H}} \times \mathbf{r}_k]^2 + U_{\text{int}} \right\} \psi. \quad (63.11)$$

We see that in certain cases (see §75) the quadratic term plays an important role.

If the magnetic field is absent and the potential energy of interaction U_{int} does not depend on the spin of the particles, for example in the case of the Coulomb interaction of charged particles, then the Hamiltonian of the system does not contain any spin variables. In this case the wave function can be written in the form of a product of the coordinate wave function and the spin wave function. The particles can be in an arbitrary spin state, while the coordinate function satisfies the ordinary Schrödinger equation.

§64. The identity of particles. The principle of identity of particles. Symmetric and antisymmetric states

We now pass on to the construction of the wave function of a system of particles of one kind, for example a system of electrons, or protons, or photons.

In such systems new important specific properties having no analogue in classical mechanics appear. These properties will become clear from the comparison of the process of collision of two particles — macroscopic and microscopic.

In classical mechanics the properties of every particle are characterized by one quantity — its mass. If the masses of two particles are the same, then the particles can be assumed to be the same. The state of each of the particles at the instant of time $t = 0$ is defined by initial conditions.

Moving in definite trajectories the particles collide elastically at a certain point of space and diverge along separate trajectories.

If the initial conditions are given, then the trajectories of each particle are completely defined and one can follow the motion of each particle. Hence particles in classical mechanics, even if they are the same, conserve their individuality. One can always establish which one of the colliding particles one has at a given point of space.

In the case of two microparticles the situation in a collision is quite different. At the instant of collision let the two particles be at definite points of space. According to the uncertainty relation, their momenta do not have sharp values. After the collision we can fix 'the trajectories' of the particles, for example two tracks in a cloud chamber. However, it is clear that if the two colliding particles are of the same kind, for example two electrons or two protons, then it is in principle impossible to establish which one of these particles is associated with a given track.

As a second example let us consider a system consisting of two hydrogen atoms.

If the atoms are at sufficiently large distances from each other, so that the electron clouds do not overlap, then each of the electrons is in effect localized by its nucleus. As the atoms approach, an overlapping of the electron clouds occurs. This means that in the region of overlap there is a certain probability of finding both electrons.

As a result of a measurement let an electron be observed in this region. It is clear that there is no way which would make it possible to establish which one of the electrons, that which earlier belonged to nucleus no. 1 or that which earlier belonged to nucleus no. 2, is observed.

The examples given above show that the 'sameness' of quantum particles has a much more profound nature than the 'sameness' of classical particles. Quantum particles are not simply the same, but are completely identical.

If we were able to change the initial state of the system by replacing each electron by the other, no physical changes would occur in the system, and this replacement could not be observed by any physical experiment.

It should also be stressed that in the examples given we have somewhat idealized the situation. If, for example, the two colliding particles have definite values of their momenta, they have no definite values of their coordinates. Hence one cannot indicate the region of collision.

Thus we arrive at the principle of identity of particles, which can be formulated in the following way: in a system of identical particles only states which do not change under the exchange of any two identical particles can be realized.

The identity of microparticles leads to very important and profound consequences. We recall that we have already encountered this property of microparticles in statistical physics. We shall now discuss the effect of the identity of particles on their collective properties in a more consistent way.

We consider a system consisting of N identical particles. The wave function of this system ψ will have the form

$$\psi(\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_k, \dots, \xi_N, t).$$

Here ξ_i is understood to be the whole set of the coordinates and spin variables characterizing the i th particle.

If two particles are exchanged, i.e. if the coordinates and the spin of the i th particle are replaced by the corresponding values for the k th particle and vice versa, then by virtue of the principle of identity the state of the system cannot change. Consequently, the wave function of the system can change only by an immaterial phase factor.

After the exchange of two particles the wave function can be expressed in terms of the initial wave function by the relation

$$\psi(\xi_1, \xi_2, \dots, \xi_k, \dots, \xi_i, \dots, \xi_N, t) = e^{i\alpha} \psi(\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_k, \dots, \xi_N, t), \quad (64.1)$$

where α is a real quantity. If the operation of exchange of the i th and k th particles is carried out once more, then the system will come back to its initial state.

On the other hand, repeating operation (64.1), we can write that

$$\begin{aligned} \psi(\xi_1, \xi_2, \dots, \xi_k, \dots, \xi_i, \dots, \xi_N, t) &= e^{i\alpha} \psi(\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_k, \dots, \xi_N, t) = \\ &= e^{2i\alpha} \psi(\xi_1, \dots, \xi_k, \dots, \xi_i, \dots, \xi_N, t). \end{aligned}$$

Hence it follows that $e^{2i\alpha} = 1$ and

$$e^{i\alpha} = \pm 1.$$

Thus when two identical particles are exchanged the wave function of the system can either remain unchanged, or change its sign. Wave functions of the first type are called symmetric, while those of the second type are called antisymmetric. We introduce the particle exchange operator \hat{P}_{ik} , which is important in what follows. By definition, when the operator \hat{P}_{ik} acts on the wave function of a system of particles $\psi(\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_k, \dots, \xi_N, t)$ it transforms it into the new function

$$\hat{P}_{ik} \psi(\xi_1, \dots, \xi_i, \dots, \xi_k, \dots, \xi_N, t) = \psi(\xi_1, \dots, \xi_k, \dots, \xi_i, \dots, \xi_N, t). \quad (64.2)$$

There corresponds to this transformation a transition of the i th particle into the state which had previously been occupied by the k th particle, and a transition of the k th particle into the state which had previously been occupied by the i th particle.

Comparing (64.2) with (64.1) we see that the eigenvalues of the operator \hat{P}_{ik} are equal to $e^{i\alpha} = \pm 1$. The symmetric and antisymmetric functions are the eigenfunctions of the operator \hat{P}_{ik} corresponding to the eigenvalues $+1$ and -1 respectively.

We shall show by means of the exchange operator that the properties of symmetry are conserved in time. This means that if at the initial instant of time the system was in either the symmetric or antisymmetric state, then no subsequent action whatever will change the character of its symmetry. In other words, the system will always remain either in the symmetric state or in the antisymmetric state. For the proof of this statement it is necessary to show that the operator \hat{P}_{ik} commutes with the Hamiltonian operator. The exchange of two identical particles corresponds only to an exchange of terms in the sum forming the Hamiltonian of the system. This is easily seen in the example of a system consisting of two identical particles. In this case the Hamiltonian can be written in the form

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla_1^2 - \frac{\hbar^2}{2m} \nabla_2^2 + U(\xi_1, t) + U(\xi_2, t) + U_{12}(\xi_1, \xi_2, t). \quad (64.3)$$

Here $U_{12}(\xi_1, \xi_2, t)$ is the energy of interaction of the particles and U corresponds to the interaction with the external field and has, obviously, the same form for the two identical particles. When the particles are exchanged, we have for the new Hamiltonian

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla_2^2 - \frac{\hbar^2}{2m} \nabla_1^2 + U(\xi_2, t) + U(\xi_1, t) + U_{12}(\xi_2, \xi_1, t). \quad (64.4)$$

It is clear that this is the same Hamiltonian as before the exchange. The result obtained is easily applied to the case of a system of N particles. We see that the exchange of particles does not change the Hamiltonian. Hence we obtain

$$\hat{H}\hat{P}_{ik} - \hat{P}_{ik}\hat{H} = 0. \quad (64.5)$$

Consequently, the symmetry properties of the system are a constant of motion and are conserved in time.

Thus it is natural to think that the symmetry is determined by the properties of the elementary particles which make up the system. Pauli managed to show that particles having an integer spin are described by symmetric functions, whereas particles having a half-integer spin are described by antisymmetric functions. The first particles are called Bose—Einstein particles, or bosons, while the latter are called Fermi—Dirac particles, or fermions. Examples of the first group of particles are light quanta (see ch. 12) and π -mesons. The second group of particles includes neutrons, protons, positrons, electrons, neutrinos and muons (all having spin $\frac{1}{2}$).

To elucidate the problem of the symmetry properties of a system consisting of identical complex particles, it is necessary to determine the total spin of the complex particle. As in the case of elementary particles, when the complex particle has an integer spin the wave function is symmetric under the exchange of the complex particles, and is antisymmetric when the complex particle has a half-integer spin.

As an example let us consider a system consisting of α -particles. For the determination of the symmetry properties of the wave function of the system it is necessary to calculate the total spin of the α -particle. The α -particle consists of two neutrons and two protons. Since the spins of the particles constituting it are equal to $\frac{1}{2}\hbar$ and the number of the particles is even, then the total spin of the α -particle is equal to an integer multiple of \hbar . Thus the wave function of a system of α -particles is a symmetric wave function.

§ 65. Wave functions for systems of fermions and bosons. The Pauli principle

Let us consider a system consisting of N non-interacting identical particles. The Schrödinger equation for the stationary states of such a system has the form

$$\sum_{i=1}^N \left[-\frac{\hbar^2}{2m} \nabla_i^2 + U(\xi_i) \right] \psi(\xi_1, \xi_2, \dots, \xi_N) = E\psi(\xi_1, \xi_2, \dots, \xi_N). \quad (65.1)$$

In §14 it was shown that the solution of this equation is the function

$$\psi = \psi_{k_1}(\xi_1)\psi_{k_2}(\xi_2) \dots \psi_{k_N}(\xi_N). \quad (65.2)$$

Here k_1, k_2, \dots are the quantum numbers of the states available to the particles. Each k_i represents the complete set of quantum numbers which characterize the state of an individual particle. The function ψ_{k_i} is the solution of the Schrödinger equation for one particle

$$-\frac{\hbar^2}{2m} \nabla_i^2 \psi_{k_i}(\xi_i) + U(\xi_i)\psi_{k_i}(\xi_i) = E_{k_i}\psi_{k_i}(\xi_i).$$

However, function (65.2) does not satisfy the requirements of symmetry. In general it is neither a symmetric nor an antisymmetric function. Since eq. (65.1) is linear, then a superposition of solutions of the type (65.2) will also be a solution. To obtain a wave function possessing the required symmetry one has to take the corresponding superposition of wave functions.

For simplicity we consider a system consisting of only two non-interacting particles. It is obvious that the functions

$$\psi_1(\xi_1, \xi_2) = \psi_1(\xi_1)\psi_2(\xi_2), \quad \psi_2(\xi_1, \xi_2) = \psi_2(\xi_1)\psi_1(\xi_2),$$

where the subscripts 1 and 2 of the wave functions $\psi_1(\xi_1)$, $\psi_1(\xi_2)$ and $\psi_2(\xi_1)$, $\psi_2(\xi_2)$ denote two different states of a particle. The wave functions $\psi_1(\xi_1, \xi_2)$, $\psi_2(\xi_1, \xi_2)$ correspond to one and the same energy of the system. From these functions one can make up two symmetrized combinations corresponding to the same energy:

$$\psi_s = C_1 [\psi_1(\xi_1)\psi_2(\xi_2) + \psi_2(\xi_1)\psi_1(\xi_2)],$$

$$\psi_a = C_2 [\psi_1(\xi_1)\psi_2(\xi_2) - \psi_2(\xi_1)\psi_1(\xi_2)].$$

The first wave function is symmetric with respect to exchange of the particles, while the second is antisymmetric. The constants C_1 and C_2 can be determined from the normalization condition. If the functions $\psi_1(\xi_1)$ and $\psi_2(\xi_2)$ are normalized to unity, while ψ_s (and ψ_a) is normalized by the condition $\int |\psi_s|^2 dV_1 dV_2 = 1$, a simple calculation gives in both cases

$$C_1 = C_2 = \frac{1}{\sqrt{2}}.$$

Hence the normalized and symmetrized function can be written in the form

$$\psi_s = \frac{1}{\sqrt{2}} [\psi_1(\xi_1)\psi_2(\xi_2) + \psi_2(\xi_1)\psi_1(\xi_2)], \quad (65.3)$$

$$\psi_a = \frac{1}{\sqrt{2}} [\psi_1(\xi_1)\psi_2(\xi_2) - \psi_2(\xi_1)\psi_1(\xi_2)]. \quad (65.4)$$

It is now easy to generalize formulae (65.3) and (65.4) to the case of an arbitrary number of non-interacting particles. Namely, for a system described by symmetric functions one can write that

$$\psi_s = \left(\frac{N!}{n_1! n_2! \dots} \right)^{-\frac{1}{2}} \sum_p \psi_{k_1}(\xi_1) \psi_{k_2}(\xi_2) \dots \psi_{k_N}(\xi_N). \quad (65.5)$$

Here n_k is the number of indices which have the same i -value. Thus the n_k show the number of particles in a given ψ_k -state. Evidently $\sum n_k = N$.

The wave function $\psi_k(\xi_k)$ is normalized to unity, all functions $\psi_k(\xi_k)$ are orthogonal to each other. Therefore in the normalization condition $\int |\psi_s|^2 dV$ only terms of the type $\int |\psi_k(\xi)|^2 d\xi$ contribute. Thus

$$\int |\psi_s|^2 dV = \sum \int |\psi_k(\xi)|^2 d\xi = \frac{N!}{n_1! n_2! \dots}.$$

Here $N!/(n_1! n_2! \dots)$ is equal to the number of rearrangements of different indices n_i . Similarly for particles with antisymmetric functions

$$\psi_a = (N!)^{-\frac{1}{2}} \begin{vmatrix} \psi_{k_1}(\xi_1) & \dots & \psi_{k_1}(\xi_N) \\ \psi_{k_2}(\xi_1) & \dots & \psi_{k_2}(\xi_N) \\ \dots & \dots & \dots \\ \psi_{k_N}(\xi_1) & \dots & \psi_{k_N}(\xi_N) \end{vmatrix} \quad (65.6)$$

The symmetrized normalized wave functions ψ_s and ψ_a describe the state of systems of N non-interacting bosons and N non-interacting fermions respectively.

Let us now consider the change in the wave function of the system when there is an interaction between the identical particles. We assume that the interaction depends on time. The exact wave function can be written in the form of one of the superpositions

$$\psi = \sum_i c_i(t) (\psi_s)_i, \quad \psi = \sum_k c_k(t) (\psi_a)_k.$$

The coefficients c_i and c_k represent the time-dependent amplitudes of the probability of the corresponding i th and k th symmetric and antisymmetric states.

The interaction gives rise to transitions in the system. As follows directly from the symmetry conservation law described in the preceding section, the

system will go over into states with the same symmetry under any external action.

Thus the wave function describing a system of interacting particles is expressed in terms of the wave functions of a system of non-interacting particles with definite symmetry.

The wave functions (65.5) and (65.6) found above make it possible to obtain a number of very important results.

Let us first of all consider a system of fermions. We assume that two particles in the system are in one and the same quantum state, i.e. that $k_1 = k_2$. This means that the two particles have the same complete set of quantum numbers, for example one and the same value of the quantum numbers n, l, m, s_z for motion in a centrally symmetric field, or p_x, p_y, p_z, s_z for free motion with a definite momentum.

Then in the determinant of (65.6) two rows turn out to be the same, and the wave function reduces identically to zero. This proves the following statement: In a system of identical fermions one cannot simultaneously have two or more particles in one and the same quantum state. This is the well-known Pauli principle, established by Pauli before the appearance of quantum mechanics on the basis of an analysis of experimental data.

The Pauli principle is often conveniently formulated in terms of the quasi-classical approximation: 'No more than one particle with given spin orientation can be found in each cell of phase space of volume $(2\pi\hbar)^3$.'

As we have seen in statistical physics, the Pauli principle determines the statistical behaviour of systems made up of identical particles with half-integer spin. The Pauli principle is of no less importance in understanding the regularities of the structure of many-electron atoms and of complex nuclei to which the next chapter is devoted.

For what follows we consider the following problem. Let a system consist of N identical particles (bosons). At a given instant of time each of the bosons is in one and the same state with the wave function $\psi(\xi)$ which is normalized in the following way:

$$\int \psi^*(\xi)\psi(\xi)dV = N.$$

Let us determine the mean energy of the system in this state. We write the Hamiltonian of such a system of particles in the form

$$\hat{H} = \sum_{i=1}^N \hat{H}_i(\xi_i) + \frac{1}{2} \sum_i \sum_j \hat{W}_{ij}(\xi_i, \xi_j), \quad (65.7)$$

where \hat{H}_i is the energy operator of the i th boson, and \hat{W}_{ij} is the operator

corresponding to the energy of interaction of the i th and j th bosons. The wave function of the system of bosons, normalized to unity, at this instant of time has the form

$$\psi(\xi_1, \xi_2, \dots, \xi_N) = N^{-\frac{1}{2}N} \psi(\xi_1) \psi(\xi_2) \dots \psi(\xi_N).$$

The mean energy of the system in this state is equal to

$$\bar{H} = \int \psi^*(\xi_1, \xi_2, \dots, \xi_N) \hat{H} \psi(\xi_1, \xi_2, \dots, \xi_N) dV_1 dV_2 \dots dV_N.$$

Taking into account the identity of bosons and assuming $N \gg 1$, we obtain

$$\bar{H} = \int \psi^*(\xi_i) \hat{H}_i \psi(\xi_i) dV_i + \frac{1}{2} \int \psi^*(\xi_i) \psi^*(\xi_j) \hat{W}_{ij}(\xi_i, \xi_j) \psi(\xi_i) \psi(\xi_j) dV_i dV_j. \quad (65.8)$$

If the particles do not interact, then $\hat{W} \equiv 0$, and the mean value of the energy has the form

$$\bar{H} = \int \psi^*(\xi_i) \hat{H}_i \psi(\xi_i) dV_i. \quad (65.9)$$

We shall need (65.8) and (65.9) for what follows.

§66. The wave function of a system of two identical particles with spin $\frac{1}{2}$

Bearing in mind further applications, let us consider the wave function of a system consisting of two particles with spin $\frac{1}{2}$, for example two electrons or two protons, in more detail.

The total wave function $\psi_n(\mathbf{r}_1, s_{1z}, \mathbf{r}_2, s_{2z})$ depends on the spatial and spin coordinates of the two particles and is antisymmetric in these variables. Assuming that there is no external magnetic field and that the interaction between the particles does not depend on their spins, we write the total wave function in the form of a product of wave functions which depend only on spatial variables and spin variables

$$\psi_n(\mathbf{r}_1, s_{1z}, \mathbf{r}_2, s_{2z}) = \Phi(\mathbf{r}_1, \mathbf{r}_2) \varphi(s_{1z}, s_{2z}). \quad (66.1)$$

We write the total spin function of the system φ in the form of a product of the eigenfunctions of the operators corresponding to the square of the spin and its z -component for each of the particles, i.e. the functions $\varphi_{\frac{1}{2}}(1)$, $\varphi_{-\frac{1}{2}}(1)$, $\varphi_{\frac{1}{2}}(2)$, $\varphi_{-\frac{1}{2}}(2)$, where the subscript denotes the spin projection of the z -axis, and the number in parentheses corresponds to the particle.

The function φ can be written in the most general form as follows:

$$\begin{aligned}\varphi(1,2) = & c_1 \varphi_{\frac{1}{2}}(1) \varphi_{\frac{1}{2}}(2) + c_2 \varphi_{-\frac{1}{2}}(1) \varphi_{-\frac{1}{2}}(2) + \\ & + c_3 \varphi_{\frac{1}{2}}(1) \varphi_{-\frac{1}{2}}(2) + c_4 \varphi_{-\frac{1}{2}}(1) \varphi_{\frac{1}{2}}(2),\end{aligned}\quad (66.2)$$

where c_1, c_2, c_3 and c_4 are arbitrary amplitudes.

We determine the spin wave functions describing the states of the system with given total spin and given z -component of spin.

Since spins are added according to the general rules of addition of angular momenta (see §52), the total spin of a system of two particles has two possible values, $S = 1$ and $S = 0$. Its z -component correspondingly has the values 1, 0 and -1 for $S = 1$, and $S_z = 0$ for $S = 0$ (in units of \hbar).

The functions φ describing a state with given S and S_z satisfy the equations

$$\hat{S}^2 \varphi = \hbar^2 S(S+1) \varphi, \quad \hat{S}_z \varphi = \hbar S_z \varphi, \quad (66.3)$$

where $\hat{S} = \hat{S}_1 + \hat{S}_2$ is the total spin operator of the system. The coefficients c_1, c_2, c_3 and c_4 in the spin function of eq. (66.2) must be chosen in a way such that both equations (66.3) for φ are automatically satisfied.

One can easily convince oneself by a direct check that the spin function of a system corresponding to all the conditions mentioned above can be written in the form

$$\left. \begin{aligned}\varphi_1^1 &= \varphi_{\frac{1}{2}}(1) \varphi_{\frac{1}{2}}(2), & S=1, S_z=1, \\ \varphi_0^1 &= \frac{1}{\sqrt{2}} [\varphi_{\frac{1}{2}}(1) \varphi_{-\frac{1}{2}}(2) + \varphi_{-\frac{1}{2}}(1) \varphi_{\frac{1}{2}}(2)], & S=1, S_z=0, \\ \varphi_{-1}^1 &= \varphi_{-\frac{1}{2}}(1) \varphi_{-\frac{1}{2}}(2), & S=1, S_z=-1,\end{aligned}\right\} \quad (66.4)$$

$$\varphi_0^0 = \frac{1}{\sqrt{2}} [\varphi_{\frac{1}{2}}(1) \varphi_{-\frac{1}{2}}(2) - \varphi_{-\frac{1}{2}}(1) \varphi_{\frac{1}{2}}(2)], \quad S=0, S_z=0, \quad (66.5)$$

where the superscript indicates the total spin of the two particles, and the subscript indicates the z -component of the spin. The expressions (66.4) and (66.5) follow from formula (52.3) and table 1 at the end of ch. 6 of coefficients C for $j_1 = \frac{1}{2}$.

This result can also be obtained from the relations (61.18). We note that the spin functions (66.4) do not change when the two particles are exchanged, i.e. under the exchange $1 \rightarrow 2, 2 \rightarrow 1$. Consequently, these functions are symmetric in the spins of the particles. The spin function (66.5) changes sign under such an exchange and is antisymmetric in the spins.

The spin functions (66.4) form a spin triplet. The whole set of the three components of a triplet is equivalent to the three-component spin function of a particle with spin one. The spin function (66.5) describing the state with spin zero forms a spin singlet.

Let us determine the eigenvalues of the scalar product $\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2$ in the singlet and triplet states. We shall have to deal with this product in what follows. Since

$$\hat{\mathbf{S}}^2 = (\hat{\mathbf{s}}_1 + \hat{\mathbf{s}}_2)^2 = \hat{\mathbf{s}}_1^2 + \hat{\mathbf{s}}_2^2 + 2\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2,$$

then

$$\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2 = \frac{1}{2}(\hat{\mathbf{S}}^2 - \hat{\mathbf{s}}_1^2 - \hat{\mathbf{s}}_2^2). \quad (66.6)$$

Substituting the eigenvalues of the operators $\hat{\mathbf{S}}^2$, $\hat{\mathbf{s}}_1^2$ and $\hat{\mathbf{s}}_2^2$ into the right-hand side, we have

$$(\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2) \varphi^S = \frac{1}{2} \hbar^2 (S(S+1) - \frac{3}{2}) \varphi^S. \quad (66.7)$$

For the triplet state $S = 1$, and thus

$$(\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2) \varphi^1 = \frac{1}{4} \hbar^2 \varphi^1. \quad (66.8)$$

Correspondingly for the singlet state $S = 0$ and

$$(\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2) \varphi^0 = -\frac{3}{4} \hbar^2 \varphi^0. \quad (66.9)$$

We now consider the function of spatial variables $\Phi(\mathbf{r}_1, \mathbf{r}_2)$. Since the total wave function (66.1) is antisymmetric, then the coordinate wave function will be antisymmetric in the state $S = 1$ and symmetric in the state $S = 0$. If the particles do not interact and are in states ψ_n and ψ_m , then the coordinate function has the form

$$\Phi_a(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} [\psi_n(\mathbf{r}_1) \psi_m(\mathbf{r}_2) - \psi_m(\mathbf{r}_1) \psi_n(\mathbf{r}_2)], \quad S = 1, \quad (66.10)$$

$$\Phi_s(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} [\psi_n(\mathbf{r}_1) \psi_m(\mathbf{r}_2) + \psi_m(\mathbf{r}_1) \psi_n(\mathbf{r}_2)], \quad S = 0. \quad (66.11)$$

As was shown in §14, it is convenient in the general case to use the coordinates $\mathbf{R} = \frac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)$ and $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, describing respectively the motion of the centre of mass of the system and the relative motion of the particles

$$\Phi(\mathbf{r}_1, \mathbf{r}_2) = \psi_0(\mathbf{R}) \psi(\mathbf{r}). \quad (66.12)$$

Let us find the consequences of the requirement of symmetry or antisymmetry of the wave function (66.12). We note first of all that if the potential of interaction of the particles depends only on the distance between them, $U(|\mathbf{r}_1 - \mathbf{r}_2|)$, then the orbital angular momentum associated with the relative motion of the particles is conserved in such a system (see §35). We now exchange the coordinates of the particles; $\mathbf{r}_1 \rightarrow \mathbf{r}_2$, $\mathbf{r}_2 \rightarrow \mathbf{r}_1$. The radius vector

of the centre of mass \mathbf{R} does not change under this exchange. Consequently, the wave function $\psi_0(\mathbf{R})$ does not change. The radius vector of the relative motion \mathbf{r} changes sign, $\mathbf{r} \rightarrow -\mathbf{r}$. As we have explained in §33, if the orbital angular momentum associated with the relative motion of the particles is given and is determined by the quantum number l , then the law of transformation of the function $\psi(\mathbf{r})$ under the replacement of \mathbf{r} by $-\mathbf{r}$ will be

$$\psi(-\mathbf{r}) = (-1)^l \psi(\mathbf{r}). \quad (66.13)$$

We see that in this case the coordinate wave function (66.12) transforms under the exchange of the particles $\mathbf{r}_1 \rightarrow \mathbf{r}_2$, $\mathbf{r}_2 \rightarrow \mathbf{r}_1$ according to the law

$$\Phi(\mathbf{r}_1, \mathbf{r}_2) \rightarrow (-1)^l \Phi(\mathbf{r}_1, \mathbf{r}_2). \quad (66.14)$$

It follows immediately from (66.14) that if the particle is in the triplet state $S = 1$, then the quantum number l can take on only odd values. On the contrary, the number l can take on only even values if the particle is in the single state $S = 0$.

§67. Exchange interaction and the concept of the chemical and strong nuclear interactions

The identity of quantum particles leads to a fundamental change in the concept of the interaction between particles.

We dwell first of all on a simple example which allows us to understand the essence of these changes. We assume that two identical particles with half-integer spin do not interact with each other in the classical sense of the word. This means that there are no terms in the Hamiltonian of the system describing an interaction between the particles.

Let one of the particles be in a given cell of phase space with the linear dimension of the cell $\sim d$. The obvious relation

$$(\Delta q_1 \Delta p_1)^3 \sim d^3 (\Delta p_1)^3 \sim \hbar^3,$$

holds, where $\Delta q_1 \sim d$ is the uncertainty in the coordinate of the particle, and $\Delta p_1 \sim \hbar/d$ is the uncertainty in its momentum.

According to the Pauli principle, the second particle cannot be in the same cell of phase space. Hence it must either be at a distance larger than d from the first particle, or have a momentum p_2 such that $|\mathbf{p}_2 - \mathbf{p}_1|$ exceeds Δp_1 , i.e. $|\mathbf{p}_2 - \mathbf{p}_1| > \hbar/d$. Only under this condition can it approach the first particle to within a distance smaller than d and be in a different cell of phase space.

We see that particles with parallel spins cannot approach each other unless they possess a sufficiently large difference of momentum. Such behaviour of the particles is equivalent to the appearance of a repulsive force between them. If the spins of the particles are antiparallel this reasoning is not valid, since the Pauli principle does not forbid such particles to be in the same cell of phase space.

Thus from the Pauli principle, which imposes restrictions upon the states of particles, it follows that there exists an interaction between the particles which depends on the orientation of their spins.

The interaction between bosons cannot be illustrated by such an obvious example. Nevertheless, it is clear that the requirement of the symmetrization of the wave function corresponds to a definite dependence of the energy of a system of particles on its total spin, i.e. it leads to an interaction between the particles.

We now assume that there is a certain weak interaction between two particles with spin $\frac{1}{2}$, described by the operator $\hat{H}'(r_{12})$, where r_{12} is the distance between the particles. For clarity, and keeping in mind further applications, we assume that \hat{H}' represents the Coulomb repulsion of two charges $\hat{H}'(r_{12}) = e^2/r_{12}$. Then the mean energy of interaction in the first approximation is equal to

$$E^{(1)} = \sum \int \psi_0^* \hat{H}' \psi_0 dV_1 dV_2. \quad (67.1)$$

Here ψ_0 is the normalized function of the unperturbed state, and the summation is carried out over all values of the spin variables.

Since in the zero order approximation particles are assumed to be non-interacting, then the spin wave functions and the coordinate wave functions are separable. The latter can be written in terms of the symmetrized or antisymmetrized products (66.10) and (66.11).

Substituting the value of the operator \hat{H}' and the wave functions into (67.1), we have

$$\begin{aligned} E^{(1)} &= \frac{1}{2} \int \frac{e^2}{r_{12}} |\psi_{n_1}(1)\psi_{n_2}(2) \pm \psi_{n_1}(2)\psi_{n_2}(1)|^2 dV_1 dV_2 = \\ &= e^2 \int \frac{|\psi_{n_1}(1)|^2 |\psi_{n_2}(2)|^2}{r_{12}} dV_1 dV_2 \pm \\ &\quad \pm e^2 \int \frac{\psi_{n_1}^*(1)\psi_{n_2}^*(2)\psi_{n_1}(2)\psi_{n_2}(1)}{r_{12}} dV_1 dV_2, \end{aligned}$$

where the numerals 1 and 2 denote respectively the coordinates of the first and second electrons, and r_{12} is the distance between them. The signs + and - refer to the states of the particles which are respectively symmetric and antisymmetric with respect to exchange of the particles. In this formula the summation has been carried out over the spin variables, which give unity. Furthermore, we have made use of the obvious equality

$$\begin{aligned} \int \psi_{n_1}^*(1) \psi_{n_2}^*(2) \frac{e^2}{r_{12}} \psi_{n_1}(2) \psi_{n_2}(1) dV_1 dV_2 &= \\ = \int \psi_{n_1}(1) \psi_{n_2}(2) \frac{e^2}{r_{12}} \psi_{n_1}^*(2) \psi_{n_2}^*(1) dV_1 dV_2. \end{aligned}$$

In this equality one integral goes over into the other when the integration indices 1 and 2 are interchanged.

Introducing the notation

$$\begin{aligned} C &= \int |\psi_{n_1}(1)|^2 \frac{e^2}{r_{12}} |\psi_{n_2}(2)|^2 dV_1 dV_2 = \\ &= \int |\psi_{n_1}(1)|^2 \hat{H}'(r_{12}) |\psi_{n_2}(2)|^2 dV_1 dV_2, \end{aligned} \quad (67.2)$$

$$\begin{aligned} A &= \int \psi_{n_1}^*(1) \psi_{n_2}^*(2) \frac{e^2}{r_{12}} \psi_{n_1}(2) \psi_{n_2}(1) dV_1 dV_2 = \\ &= \int \{\psi_{n_1}^*(1) \psi_{n_2}^*(2) \hat{H}'(r_{12}) \psi_{n_1}(2) \psi_{n_2}(1)\} dV_1 dV_2, \end{aligned} \quad (67.3)$$

we write the energy of interaction (67.1) in the form

$$E_{\uparrow\downarrow}^{(1)} = C + A, \quad (67.4)$$

$$E_{\uparrow\uparrow}^{(1)} = C - A. \quad (67.5)$$

The sign $\uparrow\downarrow$ denotes antiparallel spins (spin singlet), while the sign $\uparrow\uparrow$ denotes parallel spins (spin triplet).

It is clear from the derivation that the general form of formulae (67.4) and (67.5) obtained above is not specific, that is, it does not refer only to the case of the Coulomb interaction, but could be obtained for any interaction which depends on the coordinates of the particles.

It is interesting to compare this result with an analogous calculation for two different kinds of particles. We would then write for the non-symmetrized wave function $\Psi = \psi_{n_1}(1) \psi_{n_2}(2)$ and would correspondingly obtain

$$E' = \int |\psi_{n_1}|^2 |\psi_{n_2}|^2 \frac{e^2}{r_{12}} dV_1 dV_2. \quad (67.6)$$

Formula (67.6) has a simple meaning. It represents the mean value of the energy of the Coulomb repulsion of two particles. The position of one of the particles in the state n_1 is characterized by the probability density $|\psi_{n_1}(1)|^2$, while the position of the other particles is characterized by $|\psi_{n_2}(2)|^2$.

In formulae (67.4) and (67.5) the integral C has a structure analogous to (67.6) and is often called the Coulomb integral. However, strictly speaking, it does not allow such an interpretation, since it cannot be indicated which one of the identical particles is in the state n_1 and which one is in the state n_2 .

The integral A , usually called the exchange integral (in German Austausch, which means exchange) has no classical analogues. Calculations carried out for actual systems show that the integrals C and A are always positive. It follows directly from formulae (67.4) and (67.5) that the correction to the mean energy, determined by the interaction of the particles, depends on the orientation of their spins.

First of all we stress that it would be incorrect to assume that the interaction is made up of two parts, a classical part and an exchange part, as is often done for the sake of clarity. The contribution to the energy determined by the Coulomb integral C is called the classical part of the interaction, while the corresponding contribution of the exchange integral A is called the exchange part. In reality it is impossible to divide the interaction into the two parts, since the quantity A itself does not allow a classical interpretation.

The most characteristic part of the exchange interaction is expressed by the integral A (67.3). This integral can be treated as the matrix element corresponding to the transition of the first particle from the state n_2 into the state n_1 , and to the transition of the second particle from the state n_1 to the state n_2 . Indeed, we introduce the operator \hat{P}_{12} defined by formula (64.2), which carries out the exchange of particles, so that

$$\hat{P}_{12} \psi_{n_1}(1) \psi_{n_2}(2) = \psi_{n_1}(2) \psi_{n_2}(1),$$

$$\hat{P}_{12} \psi_{n_2}(1) \psi_{n_1}(2) = \psi_{n_1}(1) \psi_{n_2}(2).$$

Thus the operator \hat{P}_{12} is the exchange operator of the first and second particles. By means of this operator the integral A can be written as

$$\begin{aligned} A &= \int \psi_{n_1}^*(1) \psi_{n_2}^*(2) \hat{H}_{12} \psi_{n_1}(2) \psi_{n_2}(1) dV_1 dV_2 = \\ &= \int \psi_{n_1}^*(1) \psi_{n_2}^*(2) \hat{H}_{12} \hat{P}_{12} \psi_{n_1}(1) \psi_{n_2}(2) dV_1 dV_2. \end{aligned} \quad (67.7)$$

Thus the exchange interaction corresponds to the replacement of the operator \hat{H}_{12} by the operator $\hat{H}_{12}\hat{P}_{12}$. The total energy of interaction can be written by means of the exchange operator in the form

$$E' = C \pm A = \int \psi_{n_1}^*(1) \psi_{n_2}^*(2) (\hat{H}_{12} \pm \hat{H}_{12}\hat{P}_{12}) \psi_{n_1}(1) \psi_{n_2}(2) dV_1 dV_2. \quad (67.7')$$

We see that the identity of quantum-mechanical particles essentially changes their interaction. If particles of different kinds possess an arbitrary interaction characterized by the operator \hat{H}_{12} , then in the case of identical particles the operator of this interaction must be replaced by $\hat{H}_{12} \pm \hat{H}_{12}\hat{P}_{12}$. This inference does not depend on the nature of the interaction, i.e. on the character of the operator \hat{H}_{12} . Thus the electrical interaction of two identical particles (for example, two positrons) differs from the electrical interaction of different particles (for example, a positron and a proton).

Thus the fact that one has identical particles whose state is characterized by symmetrized wave functions leads to a very important general consequence: the state of the system turns out to be dependent on the total spin of the system.

This fact is a quantitative expression of the qualitative considerations presented in the beginning of this section.

The dependence of the energy of a system of particles on the total spin is equivalent to the statement of the existence of an interaction between the particles. This interaction is called the exchange interaction.

The exchange interaction has a specific quantum character. This follows formally from the fact that in the classical limit the spin of the system reduces to zero (see §60). Hence in the transition to the classical limit any difference between states with different spin, in particular the difference in their energies, vanishes.

It should be stressed that although up to now we have dealt with particles with half-integer spin, the quantitative conclusion is equally applicable to particles with integer spin, i.e. bosons. In a system of two bosons having spin zero not all states obtained as a result of the formal solution of the corresponding Schrödinger equation are realized. Only states for which the wave function is symmetric in the particles correspond to physical states of a system and to definite values of its energy. In the case of two bosons with spin 1 the energy of the system also turns out to be dependent on the total spin.

The results obtained for a system consisting of two particles, fermions or bosons, are directly applicable to the general case of a system with an arbitrary number of identical particles.

Coming back to the example of the interaction of two electrons, we shall

show that exchange forces allow the following obvious, although not rigorous, interpretation. We assume that at the instant of time $t = 0$ the first electron was in the state n_1 and the second in the state n_2 . It should be stressed once more that in reality such a formulation refers to the instant of time $t = 0$ and that the reasoning to come serves only to clarify the effect of the exchange interaction. Then the initial wave function has the form

$$\Phi(0) \equiv \frac{1}{\sqrt{2}} [\psi_a(t=0) + \psi_s(t=0)] = \psi_{n_1}(1)\psi_{n_2}(2).$$

The states described by the symmetric ψ_s and antisymmetric ψ_a wave functions are stationary states with energies respectively

$$E_s = E + C + A,$$

$$E_a = E + C - A.$$

Hence the dependence of the wave functions ψ_s and ψ_a on time is given by the formulae

$$\psi_s = \psi_s(0) e^{(-i/\hbar)(E+C+A)t},$$

$$\psi_a = \psi_a(0) e^{(-i/\hbar)(E+C-A)t}.$$

The total wave function $\Phi(t)$ for $t > 0$ is their superposition and, consequently, does not describe a stationary state.

$$\begin{aligned} \Phi(t) &= \frac{1}{\sqrt{2}} (\psi_s(t) + \psi_a(t)) = \\ &= \frac{1}{2} \{ [\psi_{n_1}(1)\psi_{n_2}(2) + \psi_{n_2}(1)\psi_{n_1}(2)] e^{(-i/\hbar)(E+C+A)t} + \\ &\quad + [\psi_{n_1}(1)\psi_{n_2}(2) - \psi_{n_2}(1)\psi_{n_1}(2)] e^{(-i/\hbar)(E+C-A)t} \} = \\ &= \{ \psi_{n_1}(1)\psi_{n_2}(2) \cos \hbar^{-1} A t - i \psi_{n_2}(1)\psi_{n_1}(2) \sin \hbar^{-1} A t \} e^{(-i/\hbar)(E+C)t}. \end{aligned} \quad (67.8)$$

Formula (67.8) shows that if at the instant of time $t = 0$ electron 1 was in the state n_1 and electron 2 in the state n_2 , then after the lapse of a time interval

$$\tau = \frac{\pi \hbar}{2A} \quad (67.9)$$

the electrons exchange states. The wave function

$$-i \psi_{n_2}(1)\psi_{n_1}(2) e^{(-i/\hbar)(E+C)t}$$

corresponds to the first electron being in the state n_2 and the second electron

in the state n_1 . After the lapse of a time interval 2τ they come back into the initial states and so on. Thus the electrons exchange states with a period τ .

Such an exchange of states is often presented in a concrete way as follows. One of the electrons of a system, for example an atom, is emitted and is then absorbed by another atom. The latter, in its turn, emits an electron which is absorbed by the first atom. In the process of 'emission' and 'capture' of the electrons a change in the momentum of the corresponding atoms occurs. The change in the momentum of the atoms means that a certain interaction between them exists. This schematic and obvious consideration of the exchange interaction justifies the term 'exchange'. However, it should not be taken literally.

This is seen particularly clearly from the following reasoning. Let the states n_1 and n_2 correspond to the bound states of the electrons in two atoms. If we tried to understand the process described above literally, in the classical sense, then a contradiction would arise. Indeed, the electrons would not manage to exchange states or be 'emitted' and 'captured' by the atoms, because for this they would need to obtain from outside a certain energy in excess of their binding energy in the atoms. In reality each of the two atoms between which there is an exchange interaction is not in a state with definite energy. The uncertainty in the energy of the system ΔE is of the order of magnitude of $\Delta E \sim A$. It makes no sense to speak of the constancy of the energy during the time interval τ , the exchange time which is in order of magnitude equal to $\Delta E \Delta t \sim A \tau \sim \hbar$. During the time τ the system is not in a state with definite energy and momentum. The two electrons are in a state with the wave function $\Phi(t)$.

In this connection it is clear that it would be inadmissible to indicate the direction of the momentum of recoil of the atoms in the 'transfer' of the electrons and to try, proceeding from this, to determine the sign of the energy of interaction. Thus when speaking of the exchange of particles it should be kept in mind that this exchange has a virtual and not a real character. The word virtual means that only the initial and final states of the system have a direct meaning.

In order that the exchange integral A may have values different from zero the wave functions ψ_{n_1} and ψ_{n_2} must overlap sufficiently, i.e. must both be different from zero in one and the same region of space. If, on the contrary, the wave functions ψ_{n_1} and ψ_{n_2} are different from zero only in different regions of space, then the exchange integral reduces to zero. If, in particular, ψ_{n_1} and ψ_{n_2} are the wave functions describing the bound states of electrons in different atoms, then the exchange interaction is possible only when the atoms are in direct contact. Further, let the wave function correspond to two

bound states in an atom. For example, n_1 is the normal state, and n_2 is one of the excited states. Then the value of the exchange integral decreases extremely rapidly in the transition to higher excited states, when the states n_1 and n_2 possess substantially different energies. Finally, when one speaks of the interaction of free particles which are described by plane waves, then the exchange integral differs from zero only for particles having similar values of the momenta. If, for example, the momenta of the particles differ appreciably, and the energy of interaction changes relatively slowly with coordinates, then A under the integral contains the product of a smoothly varying function and a rapidly oscillating function. The entire integral is then small. Thus an appreciable exchange interaction can take place only for identical particles which are in similar states, i.e. which are localized in a small region of space or have similar values of the energy and momentum.

The following important consequence results from this property of the exchange interaction; the exchange interaction possesses the property of saturation, so that in a system of a large number N of identical particles the total energy of the exchange interaction is proportional to the number of particles N . Indeed, two particles connected by the exchange interaction, for example two electrons with antiparallel spins, cannot themselves interact with a third particle.

If the ordinary energy of a pair interaction is proportional to the number of pairs, i.e. to $\frac{1}{2}N(N-1)$, then not all pairs are involved in the exchange interaction but only those which contain particles in 'similar' states (in the sense specified above). Hence the total number of particles connected by the exchange interaction is equal to the number of pairs made up of particles in similar states. It is obvious that this number of pairs is equal to $\frac{1}{2}N$.

In conclusion we point out the following fact. The derivation of the formulae for the energy of interaction was carried out with the assumption that the interaction operator does not contain any quantities which depend on the spin of the particles. However, one can also arrive at the same results in the case where the interaction operator contains spin operators.

The exchange interaction between identical particles plays a very important role in nature.

It is sufficient to point out that an exchange character is possessed by the forces to which homopolar chemical binding is due, the interaction which is responsible for the formation of crystals, the phenomenon of ferromagnetism and, finally, the interaction between particles in atomic nuclei, i.e. nuclear forces. We shall come back to the problem of chemical binding in §79, and in the meantime we shall dwell briefly on the problem of nuclear forces.

Up to now it has not been possible to construct a consistent theory of nuclear forces. The development of this theory is one of the major tasks of contemporary theoretical physics. At present the theory of nuclear forces has a semi-empirical character and is based on a number of experimental data. The totality of the available data has made it possible to establish the following properties of the nuclear interaction:

1. Experiments on the scattering of neutrons on protons show that very strong attractive forces exist between nuclear particles at distances from 1×10^{-13} cm to 2×10^{-13} cm. These forces decrease very rapidly with increasing distance and are not appreciable at distances larger than 2×10^{-13} cm. At very small distances, smaller than 1×10^{-13} cm, the attraction is replaced by a repulsion.

2. Nuclear forces turn out to be independent of the charge of the particles, i.e. the nuclear forces acting between two protons, a neutron and a proton and two neutrons are the same. The charge independence of nuclear forces follows from direct experiments on the scattering of fast neutrons and protons on protons, as well as from an analysis of the properties of the so-called mirror nuclei. Mirror nuclei are nuclei differing from each other by the interchange of neutrons and protons (nuclei with atomic numbers Z and $A - Z$, where A is the mass number, are mirror nuclei).

The identity of neutrons and protons in nuclear interactions points to a profound symmetry existing between these particles. The inequality of the masses and the presence of an electric charge on the proton are facts of relatively minor importance. Hence, according to the modern point of view, the proton and neutron should be considered as different charge states of one particle — the nucleon.

The nucleon has spin $\frac{1}{2}$ and in a given charge state obeys the Pauli exclusion principle. The nuclear interaction between nucleons is called the strong interaction (see § 112 and § 130).

3. The nucleon can be in two different charge states, the proton state and the neutron state, between which transitions are possible.

In free motion the proton state, which has a smaller mass and energy, is more stable. Hence the free neutron decays according to the scheme

$$n \rightarrow p + e + \bar{\nu},$$

where $\bar{\nu}$ is the antineutrino.

In atomic nuclei, where there is a nuclear interaction between the particles, the transformation of neutrons and protons into each other occurs (see below).

4. The presence of a charge on the proton entails two consequences: (1)

the proton state and the neutron state are different states of the nucleon; (2) besides the nuclear interaction there are the forces of Coulomb repulsion between two protons. These forces become important in the case of heavy nuclei, determining their instability.

5. The nuclear interaction depends not only on the distance but also on the mutual orientation of the spins of the interacting particles, as well as on the orientation of the spins with respect to the axis joining the two nucleons.

The dependence of the nuclear interaction on the orientation of the spins follows directly from experiments on the scattering of very slow neutrons on orthohydrogen and parahydrogen.

The existence of the dependence on the orientation of the spins with respect to the axis follows from an analysis of the properties of the deuteron, in particular from its possession of a quadrupole moment.

6. The nuclear interaction has exchange character. This fundamental conclusion follows, first of all, from the very fact of the stability of nuclei.

If the nuclear (strong) interaction depended only on the distance between the particles, then the potential energy of a system with mass number A would be proportional to A^2 — to the number of pairs which attract each other. However, the kinetic energy of a gas of Fermi particles confined in a given volume increases with the number of particles, according to (79.4) of Part III, as $A^{5/3}$.

Thus for a sufficiently large mass number the potential energy would turn out to be larger than the kinetic energy, the nucleus would have to contract and the particles would have to merge with each other. The volume of the nucleus would be a constant quantity which does not depend on A , and its binding energy would be proportional to A^2 . In reality the data on scattering show that the volume of the nucleus increases in proportion to A , and that the binding energy is also proportional to A . This means that nuclear forces possess the property of saturation. Saturation, as we have seen above, is a characteristic property of exchange forces.

We shall dwell in somewhat more detail on the description of the modern concepts of the nature of nuclear forces.

The following simple picture of nuclear forces results from the assumption that the proton and neutron are different states of one particle and from the exchange character of the nuclear interaction: between two nucleons at very small distances there is a virtual exchange of a certain particle which is called the 'carrier' of the interaction. This exchange is in principle similar to the virtual exchange of the electron which was considered in detail above in the example of the exchange interaction.

It turned out (see below) that the particle responsible for the nucleon-nucleon interaction is the π -meson.

Three types of exchange are possible:

$$\left. \begin{aligned} p &\rightleftharpoons n + \pi^+, \\ n &\rightleftharpoons p + \pi^-, \\ p &\rightleftharpoons p + \pi^0, \\ n &\rightleftharpoons n + \pi^0. \end{aligned} \right\}$$

In the first two virtual processes the nucleon goes over from the proton state into the neutron state and vice versa; in the last virtual process the charge state of the nucleon does not change. The process of exchange of the charged meson can clearly be interpreted in the same way as the exchange of electrons: each of the nucleons spends a part of its time in the charged state and a part of its time in the neutral state. The exchange of π -mesons gives rise to the attraction between nucleons. We have emphasized the virtual character of the exchange, since an energy not less than $m_\pi c^2$, where m_π is the mass of the π -meson, would be necessary for the production of real π -mesons.

All π -mesons, positive, negative and neutral, should be considered as different charge states of one particle.

Further, it turns out that the masses of π -mesons cannot be arbitrary. They can be connected with the range of nuclear forces. Since nuclear forces do not depend on the electric charge and have a purely quantum nature, the range of the forces can depend only on the mass of the carrier particles m_π and the universal constants \hbar and c .

From the above three quantities one can construct only one constant with the dimensionality of length; the Compton wavelength of the meson

$$R \sim \frac{\hbar}{m_\pi c}. \quad (67.10)$$

On the basis of the following reasoning one can ascribe an obvious meaning to expression (67.10). In the virtual exchange of the meson the energy of each of the nucleons must have an uncertainty $\Delta E \sim m_\pi c^2$. The exchange time τ must have an order of magnitude of $\tau \sim \hbar/m_\pi c^2$. If it is assumed that the meson moves with a velocity $\sim c$, then during the time τ it traverses the path

$$R \sim c\tau \sim \frac{\hbar}{m_\pi c}.$$

This distance is just the range of nuclear forces. Giving the range of nuclear forces, one can find the mass of the carrier particles

$$m_{\pi} \sim \frac{\hbar}{Rc} \approx 300m_{e1},$$

where m_{e1} is the mass of the electron. This is of the same order of magnitude as the mass of the π -meson. We note that π -mesons were discovered experimentally after having been introduced in the theory as the hypothetical particles responsible for the strong nuclear interaction. The most convincing proof that π -mesons are the carriers of nuclear forces is the experimentally established extremely strong interaction of π -mesons with nucleons.

If the energy of a system of nucleons exceeds $m_{\pi}c^2$, then π -mesons can really be produced. The appearance of π -mesons was observed in collisions of fast nucleons (see §136) as well as in the action of γ -rays on nucleons. The reaction

$$p + \gamma \rightarrow n + \pi^+,$$

representing the elementary act of the nuclear photoeffect was observed. Later, in particular in §112, we shall come back to the problem of nuclear forces.

Applications of Quantum Mechanics to Atomic and Nuclear Systems

§68. The helium atom

The hydrogen atom, which was considered in detail in ch. 4, is the simplest one-electron system. Passing to the study of many-electron systems it is natural to investigate, first of all, the properties of the helium atom, in which two electrons revolve around the nucleus. We assume that the nucleus has an infinitely large mass. Hence, assuming the nucleus to be at rest, we write the Hamiltonian of the system of two electrons in the form

$$\left(-\frac{\hbar^2}{2m} \nabla_1^2 - \frac{\hbar^2}{2m} \nabla_2^2 - \frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{e^2}{r_{12}} \right) \psi = E\psi. \quad (68.1)$$

Here \mathbf{r}_1 and \mathbf{r}_2 are the radius vectors of the first and second electrons, and r_{12} is the distance between them. The third and fourth terms of (68.1) express the potential energy of the electrons in the field of the nucleus, and the last term expresses the energy of the Coulomb interaction between the electrons.

It should be noted that a number of approximations are associated with the representation of the Hamiltonian in such a form. Electrons possess a magnetic moment whose interaction is of a more complex character than the Coulomb interaction. Further, magnetic moments (the spin magnetic moment and the orbital magnetic moment) also interact with each other. However, we

shall not study these effects, which have the character of small corrections, in detail.

Since the Hamiltonian of the system does not contain any spin operators, the solution of eq. (68.1) must be sought in the form of a product of two functions, one of which depends only on the coordinates and the other only on the spin

$$\psi = \Phi(\mathbf{r}_1, \mathbf{r}_2) \varphi(s_{z1}, s_{z2}). \quad (68.2)$$

In §66 it was shown that the spin function of two electrons is symmetric with respect to the exchange of two particles if the total spin of the system is equal to one, and that it is antisymmetric if the total spin is equal to zero. Thus it is seen that the states of the helium atom are divided into two groups. States with zero spin are called parastates, and those with unity spin are called orthostates. If the Hamiltonian (68.1) described the system exactly, then the three orthostates differing in the z -component of the spin would have the same energy. However, the weak interaction between the spin magnetic moment and the orbital magnetic moment removes the degeneracy and three close sub-levels arise. Thus the energy spectrum of helium consists of a set of singlet and triplet levels.

From general considerations it is easy to establish the group of states to which the ground state of helium belongs. As is known, the ground state of helium is described by a wave function having no nodes (see §10). It is obvious that this function cannot be an antisymmetric coordinate function, since the latter reduces to zero for $\mathbf{r}_1 = \mathbf{r}_2$.

Indeed, if $\Phi(\mathbf{r}_1, \mathbf{r}_2)$ is an antisymmetric function of the two variables \mathbf{r}_1 and \mathbf{r}_2 , then it satisfies the relation

$$\Phi(\mathbf{r}_1, \mathbf{r}_2) = -\Phi(\mathbf{r}_2, \mathbf{r}_1). \quad (68.3)$$

For $\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r}$ we have $\Phi(\mathbf{r}, \mathbf{r}) = 0$.

Thus we see that in the normal state the wave function is symmetric in the coordinates and, consequently, is antisymmetric in the spins.

The normal state of helium is a parastate.

In eq. (68.1) the variables are not separable and it is impossible to obtain the exact solution. Therefore a number of approximate methods have been devised for its solution. The application of perturbation theory makes it possible to obtain the wave functions and, in a rather rough approximation, the energy of the ground state of helium.

We shall assume that the interaction between the electrons is the perturbation in eq. (68.1).

Then in the zero order approximation (68.1) can be written in the form

$$\left(-\frac{\hbar^2}{2m}\nabla_1^2 - \frac{\hbar^2}{2m}\nabla_2^2 - \frac{2e^2}{r_1} - \frac{2e^2}{r_2}\right)\Phi_0 = E_0\Phi_0. \quad (68.4)$$

This equation can be solved by the method of separation of variables. For the normal state of the helium atom we have

$$\Phi_0 = \psi_1(\mathbf{r}_1)\psi_1(\mathbf{r}_2), \quad E_0 = 2E_1,$$

where E_1 and ψ_1 denote respectively the energy and the wave function of the normal state of a hydrogen-like atom with charge $Z = 2$. As we have pointed out before, the function Φ_0 is symmetric in the coordinates of the electrons.

The energy level of the ground state in the first approximation of perturbation theory is given by the formula

$$E = E_0 + E^{(1)},$$

where the quantity $E^{(1)}$ in accordance with formula (53.12) is determined by the matrix element

$$E^{(1)} = \int |\psi_1(\mathbf{r}_1)|^2 |\psi_1(\mathbf{r}_2)|^2 \frac{e^2}{r_{12}} dV_1 dV_2. \quad (68.5)$$

In calculating higher energy levels of the helium atom the solution of the unperturbed wave equation can be written in the form

$$\Phi_0 = \psi_n(\mathbf{r}_1)\psi_m(\mathbf{r}_2), \quad E_0 = E_n + E_m. \quad (68.6)$$

Here ψ_n and ψ_m are the wave functions of the hydrogen-like atom in the n th and m th states respectively. (We shall not take the degeneracies in the orbital and magnetic quantum numbers further into account in obtaining a qualitative picture.)

It is easily understood that the function

$$\Phi_0 = \psi_n(\mathbf{r}_2)\psi_m(\mathbf{r}_1), \quad E_0 = E_n + E_m \quad (68.7)$$

will also be a solution of the unperturbed equation.

Thus a two-fold degeneracy arises. The two solutions (68.6) and (68.7) differ from each other by the exchange of electrons. For further calculations we shall need to have recourse to perturbation theory in the presence of degeneracy (see §54).

The correction to the energy E_0 in the first approximation is determined in this case from the condition that the determinant below be zero (see §54)

$$\begin{vmatrix} H'_{11} - E^{(1)} & H'_{12} \\ H'_{21} & H'_{22} - E^{(1)} \end{vmatrix} = 0.$$

The quantities H'_{11} , H'_{12} , H'_{21} and H'_{22} represent the matrix elements

$$\begin{aligned} H'_{11} &= \int \psi_n^*(\mathbf{r}_1) \psi_m^*(\mathbf{r}_2) \frac{e^2}{r_{12}} \psi_n(\mathbf{r}_1) \psi_m(\mathbf{r}_2) dV_1 dV_2, \\ H'_{22} &= \int \psi_m^*(\mathbf{r}_1) \psi_n^*(\mathbf{r}_2) \frac{e^2}{r_{12}} \psi_m(\mathbf{r}_1) \psi_n(\mathbf{r}_2) dV_1 dV_2, \\ H'_{12} &= \int \psi_n^*(\mathbf{r}_1) \psi_m^*(\mathbf{r}_2) \frac{e^2}{r_{12}} \psi_n(\mathbf{r}_2) \psi_m(\mathbf{r}_1) dV_1 dV_2, \\ H'_{21} &= \int \psi_m^*(\mathbf{r}_1) \psi_n^*(\mathbf{r}_2) \frac{e^2}{r_{12}} \psi_n(\mathbf{r}_1) \psi_m(\mathbf{r}_2) dV_1 dV_2. \end{aligned} \quad (68.8)$$

It is easily seen that $H'_{11} = H'_{22}$, $H'_{12} = H'_{21}$. Indeed, if \mathbf{r}_1 is replaced by \mathbf{r}_2 and \mathbf{r}_2 by \mathbf{r}_1 in the expression for the matrix element H'_{11} , then we obtain exactly the expression H'_{22} . An analogous statement also holds for the matrix elements H'_{12} and H'_{21} . Evaluating the determinant, we get

$$(H'_{11} - E^{(1)})^2 - H'^2_{12} = 0.$$

Hence we find two values for the correction to the energy:

$$E^{(1)}_1 = H'_{11} + H'_{12}, \quad (68.9)$$

$$E^{(1)}_2 = H'_{11} - H'_{12}. \quad (68.10)$$

The expressions (68.9) and (68.10) are the same as the general formula found in §67. To the two values of the energy (68.9) and (68.10) there correspond two wave functions of the type $\psi = a\psi_n(\mathbf{r}_1)\psi_m(\mathbf{r}_2) + b\psi_n(\mathbf{r}_2)\psi_m(\mathbf{r}_1)$. The coefficients a and b are determined from the equations

$$(H'_{11} - E^{(1)})a + H'_{12}b = 0, \quad H'_{21}a + (H'_{22} - E^{(1)})b = 0.$$

For the value of $E^{(1)}_1$ we obtain $a = b$, while in the case of $E^{(1)}_2$ we have $a = -b$. Thus under the action of the perturbation the degeneracy is removed and we obtain two different states

$$\begin{aligned} \Phi^{(s)}_0 &= a_1 [\psi_n(\mathbf{r}_1)\psi_m(\mathbf{r}_2) + \psi_n(\mathbf{r}_2)\psi_m(\mathbf{r}_1)], \\ \Phi^{(a)}_0 &= a_2 [\psi_n(\mathbf{r}_1)\psi_m(\mathbf{r}_2) - \psi_n(\mathbf{r}_2)\psi_m(\mathbf{r}_1)]. \end{aligned} \quad (68.11)$$

where a_1 and a_2 are the normalization constants.

As is seen from formulae (68.11), the states $\Phi^{(s)}_0$ and $\Phi^{(a)}_0$ are respectively symmetric and antisymmetric in the coordinates of the electrons. In correspondence with the aforesaid, the function $\Phi^{(s)}_0$ describes the state of the

helium atom with spin zero, while the wave function $\Phi_0^{(a)}$ corresponds to the state with total spin equal to one.

The wave function $\Phi_0^{(s)}$ corresponds to a parastate with zero spin and higher energy $E_1^{(1)}$. Correspondingly, $\Phi_0^{(a)}$ corresponds to an orthostate with unit spin and lower energy $E_2^{(1)}$.

The matrix elements $H'_{11} = H'_{22}$, as is seen from their definition, represent the Coulomb integral, while $H'_{12} = H'_{21}$ represent the exchange integral. All that was said in §67 applies to the system of two electrons in helium. For example, if electron 1 is in the ground state, while electron 2 is in an excited state, then after a lapse of time $\tau = \pi\hbar/2|H'_{12}|$ they exchange states.

The relations (68.9) and (68.10) still do not give a complete picture of the levels of the atom. Indeed, in the calculation carried out above we have not taken into account the degeneracy in the quantum number l of the levels of hydrogen-like atoms. The interaction between the electrons removes this degeneracy, and the levels turn out to be dependent not only on the principal quantum numbers but also on the orbital angular momentum.

We note that the method of calculation given above does not give high accuracy. The energy of the ground level of the helium atom obtained according to the above theory differs by about 20% from the value found experimentally. This strong disagreement is due to the choice of the perturbation, which is not sufficiently small.

§69. The variational principle

We have already seen that the two-electron problem, the helium atom, cannot be solved accurately and calls for the use of approximate methods.

This applies even more to complex atoms containing many electrons. Despite the complexity of many-electron atoms, effective approximate methods of solution allow one to get a very detailed idea of their properties. The effective approximate methods are to a large degree connected with the extremal properties of the Schrödinger equation. Namely, it turns out that the Schrödinger equation can be obtained from a variational principle.

We introduce the functional

$$J = \int \varphi^* \hat{H} \varphi dV, \quad (69.1)$$

where the restriction

$$\int \varphi^* \varphi dV = 1 \quad (69.2)$$

is imposed upon the function φ . In other respects φ remains an arbitrary

complex function having the same dimensionality as the eigenfunctions of the operator \hat{H} . The minimum value of the functional J under the condition (69.2) can be found by Lagrange's method. In varying the complex function it is possible to vary φ and φ^* independently. For concreteness, let us vary φ^* . The variation of φ leads to the same result.

Clearly we have

$$\delta \int \varphi^* \hat{H} \varphi dV + E_0 \delta \int \varphi^* \varphi dV = 0,$$

where E_0 is the Lagrange multiplier. Hence we have

$$\int \delta \varphi^* (\hat{H} - E_0) \varphi dV = 0, \quad (69.3)$$

or in view of the arbitrariness of $\delta \varphi^*$,

$$(\hat{H} - E) \varphi = 0. \quad (69.4)$$

Thus, if $\varphi = \psi_0$, where ψ_0 is the normalized solution of the Schrödinger equation corresponding to the eigenvalue E_0 of the operator \hat{H} , then the functional J is equal to

$$J(\psi_0) = \int \psi_0^* \hat{H} \psi_0 dV = E_0. \quad (69.5)$$

We shall show that E_0 is the minimum eigenvalue of \hat{H} , i.e. the energy of the ground state. Let $\varphi = \psi_0 + \sum c_n \psi_n$. Then for J we find

$$J = \int (\psi_0^* + \sum c_n^* \psi_n^*) \hat{H} (\psi_0 + \sum c_n \psi_n) dV = E_0 + \sum |c_n|^2 E_n \geq E_0. \quad (69.6)$$

The wave functions of the excited stationary states ψ_n must satisfy not only condition (69.2) but also the condition of orthogonality

$$\int \psi_0 \psi_n dV = 0. \quad (69.7)$$

They correspond to extrema but not the minimum of $J(\psi_n)$.

The variational properties of the Schrödinger equation are widely used for obtaining approximate solutions of it. Defining the form of a trial function on the basis of physical considerations or experimental data one seeks the minimum value of the integral $J(\varphi)$.

Let us consider as an example the harmonic oscillator. Choosing as the trial function the normalized function $\varphi = (2\alpha/\pi)^{1/4} e^{-\alpha x^2}$ we have

$$J(\varphi) = \left(\frac{2\alpha}{\pi}\right)^{1/4} \int e^{\alpha x^2} \left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{m\omega^2 x^2}{2}\right) e^{-\alpha x^2} dx = \frac{\hbar^2 \alpha}{2m} + \frac{m\omega^2}{8\alpha}.$$

The condition of minimum gives $\alpha = m\omega/2\hbar$. Hence

$$J_{\min} = E_0 = \frac{1}{2}\hbar\omega$$

and

$$\varphi_{\min} = \psi_0 = (m\omega/\hbar\pi)^{\frac{1}{2}} \exp(-m\omega x^2/2\hbar).$$

The appropriate choice of the trial function φ led us to an accurate value of E_0 and ψ_0 . Had we chosen another trial function, then we would have obtained another, although similar, value E'_0 and ψ'_0 . A shortcoming of the variational method is the fact that it gives an unpredictable error.

Other examples of the use of the variational principle will be given in the next section.

§70. The self-consistent field method (Hartree–Fock method)

For the calculation of many-electron systems wide use is made of the self-consistent field method which we have already encountered (see §41 of Part IV). The idea of the method (often called the Hartree–Fock method) is as follows. In the zero order approximation all the electrons are assumed to move independently of each other in the field of the nucleus. By means of the wave functions of the zero order approximation one finds the charge density and the mean electrostatic field produced by all the electrons.

In the next approximation each of the electrons is assumed to move in the field of the nucleus and the field produced by all the other electrons. The solution of the Schrödinger equation in this field gives the wave function in the first order approximation. Introducing the correction into the charge distribution and field distribution and solving the Schrödinger equation in the new field, one can find the correction of the second order approximation and so on.

To obtain the Schrödinger equation in the self-consistent field method we shall make use of the variational principle. To abbreviate the notation we shall carry out the calculations for the example of a two-electron system (the helium atom), confining ourselves to the calculation of the ground state. Therefore we shall not take into account the requirement of symmetrization of the wave function of the system of electrons. This will be done somewhat later. In the zero order approximation both electrons are described by identical real wave functions ψ , and the wave function of the atom has the form

$$\Psi = \psi_1 \psi_2. \quad (70.1)$$

The variational principle reads

$$\begin{aligned}\delta \int \Psi^* (\hat{H} - E) \Psi \, dV &= \delta \int \psi_1 \psi_2 (\hat{H} - E) \psi_1 \psi_2 \, dV = \\ &= \int \psi_2 \delta \psi_1 (\hat{H} - E) \psi_1 \psi_2 \, dV = 0.\end{aligned}\quad (70.2)$$

Hence

$$\int \psi_2 (\hat{H} - E) \psi_1 \psi_2 \, dV = 0. \quad (70.3)$$

Substituting the value of H from (68.1) into (70.3), we obtain

$$\left\{ -\frac{\hbar^2}{2m} \nabla_1^2 + E_1 - \frac{2e^2}{r_1} + \int \psi_2^2 \frac{e^2}{r_{12}} \, dV_2 \right\} \psi_1 = 0. \quad (70.4)$$

Here the additional term in the potential energy has the simple meaning

$$C(\mathbf{r}_1) = \int \psi_2^2 \frac{e^2}{r_{12}} \, dV_2 = \int \frac{e\rho(\mathbf{r}_2)}{r_{12}} \, dV_2, \quad (70.5)$$

where $\rho(\mathbf{r}_2)$ is the charge density produced by the second electron. The same equation is obtained for ψ_2 . The total energy of the atom turns out not to be equal to twice the value of E_1 but is given by the formula

$$E = 2E_1 - \int \frac{\rho_1 \rho_2}{r_{12}} \, dV_1 \, dV_2.$$

Indeed, by definition

$$E = \int \Psi^* \hat{H} \Psi \, dV = \int \Psi^* \left[\hat{H}_1 + \hat{H}_2 + \frac{e^2}{r_{12}} - C(\mathbf{r}_1) - C(\mathbf{r}_2) \right] \Psi \, dV = 2E_1 - \bar{C}. \quad (70.6)$$

Here

$$E_1 = E_2 = \int \Psi^* \hat{H}_1 \Psi \, dV,$$

where the operator \hat{H}_1 is equal to

$$\hat{H}_1 = -\frac{\hbar^2}{2m} \nabla_1^2 - \frac{2e^2}{r_1} + C(\mathbf{r}_1).$$

The quantity \bar{C} is equal to

$$\bar{C} = \int \frac{\rho_1 \rho_2}{r_{12}} \, dV.$$

It is obvious that \bar{C} represents the mean energy of the electrostatic interaction between the electrons. To obtain the correct value of the energy E it is

necessary to subtract \bar{C} from $2E_1$, since this quantity is involved in the Hamiltonian of each of the electrons. In the case of a system of N electrons an analogous derivation gives for the i th electron in the n th quantum state

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + U(\mathbf{r}_i) + \sum e_i e_k \int \frac{|\psi_{nk}|^2 dV_k}{|\mathbf{r}_i - \mathbf{r}_k|} \right\} \psi_{ni} = E_n \psi_{ni}. \quad (70.7)$$

The structure of the general equation does not differ from that of equation (70.4). The complexity of the Schrödinger equations in the self-consistent field approximation is associated with the fact that the equation for ψ_i involves the wave functions of all the other electrons. Therefore, even in the simplest case of a two-electron system, one has to solve eq. (70.4) either by numerical or by approximate methods; for example the variational method. In this latter case it is natural to choose as the trial functions the hydrogen-like functions for a certain effective nuclear charge. The value of this charge is found from the condition of minimum of the integral (70.2). These calculations, as well as a summary of the numerical solutions, can be found in the book of Bethe and Salpeter*.

So far we have not taken into account the symmetry of the wave function. It is clear, however, that from the theoretical point of view the symmetrization of the wave function must be carried out from the very beginning of the calculation. For example, if no account is taken of the symmetry of the wave function no difference appears in the energy of orthohelium and parahelium.

The self-consistent field method taking account of the requirements of symmetry of the wave function is called the Hartree–Fock method. In the simplest case of two electrons all the preceding calculations can easily be carried out for the symmetrized wave function

$$\Psi(1,2) = \frac{1}{\sqrt{2}} [\psi_1(1)\psi_2(2) - \psi_2(1)\psi_1(2)].$$

Substituting this expression into (70.2) we have to vary the wave functions ψ_1 and ψ_2 independently of each other.

Then instead of (70.3) we obtain

$$\sum_{i=1}^2 \iint dV_1 \delta\psi_i(1) dV_2 \{ \psi_k(2)(\hat{H} - E) [\psi_k(2)\psi_i(1) \pm \psi_i(2)\psi_k(1)] = 0. \quad (70.8)$$

* H.A. Bethe and E.E. Salpeter, *Quantum mechanics of one and two electron systems*, Handbuch der Physik, vol. 35 (Springer, Berlin, 1957).

In this case in (70.8) $i = 1$ for $k = 2$ and $i = 2$ for $k = 1$. In view of the arbitrariness of $\delta\psi_1$ and $\delta\psi_2$ we arrive at two equations. On substituting the total operator \hat{H} from (68.1), these equations assume the form

$$\left\{ -\frac{\hbar^2}{2m} \nabla_1^2 - E - \frac{2e^2}{r_{12}} + H_{22} + C_{22} \right\} \psi_1(\mathbf{r}) - [H_{12} + C_{12}] \psi_2(\mathbf{r}) = 0, \quad (70.9)$$

$$\left\{ -\frac{\hbar^2}{2m} \nabla_2^2 - E - \frac{2e^2}{r_{12}} + H_{11} + C_{11} \right\} \psi_2(\mathbf{r}) - [H_{12} + C_{12}] \psi_1(\mathbf{r}) = 0,$$

where

$$C_{12} = \int \psi_1(2) \psi_2(1) \frac{e^2}{r_{12}} dV_2, \quad H_{ik} = \int \psi_i \left(-\frac{\hbar^2}{2m} \nabla_2^2 - \frac{2e^2}{r_{12}} \right) \psi_k dV.$$

By taking into account the symmetry of the wave function the number of unknown wave functions is doubled, and a system of simultaneous equations is obtained. The main difference between the Hartree–Fock equations and the Hartree equations consists in the appearance of exchange integrals, i.e. terms of the form C_{12} .

In the general case of many-electron atoms the wave function of the system of electrons which is to be substituted into the equation of the variational principle must be written in the form (65.6). We shall not give the cumbersome equations which are then obtained. Although in solving the Hartree–Fock equations numerically one has to carry out very laborious calculations, it is possible to find with a high degree of accuracy the energy of the ground and excited states, and the distribution of the charge and of the field for helium, as well as for a number of other atoms and ions. Naturally the number of numerical calculations necessary in integrating the Hartree–Fock equations increases rapidly with increasing number of electrons.

§ 71. The statistical model of the atom

For heavy atoms, when the calculation of the many-electron system according to the Hartree–Fock method becomes very time-consuming, a statistical method is widely adopted.

Let a system of a large number of electrons move in a spherically symmetric field $\varphi(r)$. By virtue of the Pauli principle a large fraction of these electrons will be in states with large quantum numbers. If the potential $\varphi(r)$

changes sufficiently slowly in space, then the electrons can be considered in the quasi-classical approximation. If, furthermore, the interaction between the electrons is sufficiently weak, then the whole set of electrons can be considered to be an ideal Fermi gas at absolute zero.

In a degenerate Fermi gas (see §79 of Part III) the electrons occupy quantum states in pairs, so that there is a phase-space cell of volume $(2\pi\hbar)^3$ per pair. In this case, all cells in momentum space with a momentum lying in the interval $0 \leq p \leq p_{\max}$ are filled. The value p_{\max} is easily expressed in terms of the electron gas density n (i.e. in terms of the mean number of electrons per unit volume). The number of electrons per unit volume with a given value of the momentum is evidently equal to

$$dn = 2 \frac{4\pi}{(2\pi\hbar)^3} p^2 dp.$$

Integrating from $p = 0$ to $p = p_{\max}$ we have

$$p_{\max}^3 = \frac{3}{8\pi} (2\pi\hbar)^3 n. \quad (71.1)$$

This formula allows the charge density to be expressed in terms of the momentum

$$\rho = \frac{8\pi e}{3(2\pi\hbar)^3} p_{\max}^3. \quad (71.2)$$

On the other hand, p_{\max} can be related to the potential by means of the following simple reasoning. The energy of an electron bound in an atom, E , is always negative, i.e.

$$E = \frac{p^2}{2m} + e\varphi(r) \leq 0.$$

We assume that the potential $\varphi(r)$ reduces to zero outside the atom. Hence for the maximum momentum compatible with the requirement $E = 0$ we find

$$p_{\max} = [-2me\varphi(r)]^{\frac{1}{2}}.$$

Hence the electron charge density is connected with the potential by the relation

$$\rho = \frac{8\pi e(-2me)^{\frac{3}{2}} \varphi^{\frac{3}{2}}}{3(2\pi\hbar)^3}. \quad (71.4)$$

In the self-consistent field approximation one can write for the potential of the electrostatic field $\varphi(r)$ the Poisson equation

$$\nabla^2 \varphi = -4\pi\rho,$$

or, taking into account the spherical symmetry of the atom,

$$\begin{aligned} \frac{1}{r} \frac{d^2(r\varphi)}{dr^2} &= - \frac{32\pi^2 e (-2me)^{\frac{2}{3}} \varphi^{\frac{2}{3}}}{3(2\pi\hbar)^3} = \\ &= - \frac{4e(-2me)^{\frac{2}{3}} \varphi^{\frac{2}{3}}}{3\pi\hbar^3}. \end{aligned} \quad (71.5)$$

The equation obtained is called the Thomas–Fermi equation. To obtain the distribution of the potential $\varphi(r)$ it is necessary to supplement this equation with boundary conditions. Let us first consider the case of neutral atoms. Then one of the boundary conditions is $\varphi \rightarrow 0$ as $r \rightarrow \infty$. The second condition follows from the requirement that near the nucleus, when its charge is not screened by electrons, the field be a purely Coulomb field, i.e. that

$$\varphi(r) \rightarrow \frac{Ze}{r} \quad \text{as} \quad r \rightarrow 0. \quad (71.6)$$

To obtain the solution of eq. (71.5) with boundary conditions (71.3) and (71.6) it is convenient to pass to dimensionless quantities, defining them by the relations

$$\chi = \frac{r\varphi}{Z|e|}, \quad x = \frac{r}{d},$$

where d is a constant quantity with the dimensionality of length. For χ we find the equation

$$\frac{d^2\chi}{dx^2} = \frac{|e|^3 (2m)^{\frac{2}{3}} Z^{\frac{1}{3}} d^{\frac{2}{3}}}{3\pi\hbar^3} \frac{\chi^{\frac{2}{3}}}{x^{\frac{1}{3}}}.$$

Setting d equal to

$$d = \frac{1}{2} \left(\frac{9\pi^2}{16} \right)^{\frac{1}{3}} \frac{\hbar^2}{me^2} \frac{1}{Z^{\frac{1}{3}}} = 0.88 \frac{a}{Z^{\frac{1}{3}}}, \quad (71.7)$$

where a is the radius of the Bohr orbit, we arrive at the equation

$$\frac{d^2\chi}{dx^2} = \frac{\chi^{\frac{2}{3}}}{x^{\frac{1}{3}}}. \quad (71.8)$$

In this case it is obvious that

$$\begin{aligned}\chi &\rightarrow 1 & \text{as } x &\rightarrow 0, \\ \chi &\rightarrow 0 & \text{as } x &\rightarrow \infty.\end{aligned}\tag{71.9}$$

The integration of eq. (71.8) with boundary conditions (71.9) has been carried out numerically. Since the boundary value problem does not depend on the atomic number, the integration of this system allows one to find the universal distribution of the dimensionless potential in an atom.

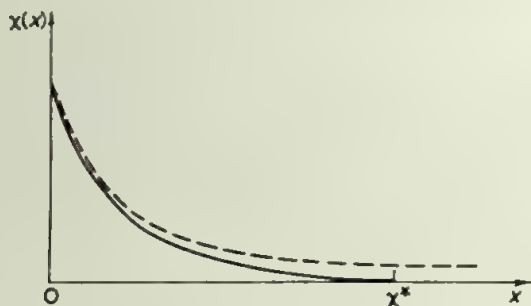


Fig. V.19

The behaviour of the function $\chi(x)$ for an atom is shown in fig. V.19 by the dotted line*. Since the function $\chi(x)$ for $x \rightarrow \infty$ only reduces to zero asymptotically, the potential, and also the electron density, nowhere reduces to zero. This means that in the approximation considered a finite value of the atomic radius cannot be found.

In fig. V.20 the curve of the radial electron density $D = 4\pi r^2 \rho(r)$ for the argon atom according to Thomas–Fermi (solid line) is compared with the result of the Hartree–Fock method (dotted line).

Fig. V.20 illustrates in an obvious way the merits and shortcomings of the Thomas–Fermi method. It does not give all the details of the behaviour of the electron density inside the atom, but it makes it possible to establish sufficiently accurately its general trend.

* The tabulated values of the function $\chi(x)$ can be found in the following books: L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965) and P.Gombas, *Die statistische Theorie des Atoms und Ihre Anwendungen* (Springer Verlag, Wien, 1949).

In the outer parts of the atom, at a large distance from the nucleus, the electron density as calculated by the Thomas–Fermi method is overestimated.

The fact that the Thomas–Fermi method gives poor results for the peripheral regions of the atom follows from the conditions of its applicability (see below). The numerical calculation of the behaviour of the electron density with distance from the nucleus shows that one half of the total electron charge is contained in a sphere of radius $R \approx 1.33aZ^{-1/3}$.

Therefore, qualitatively, the quantity R can be considered to be the effective radius of the atom. It decreases with increasing Z .

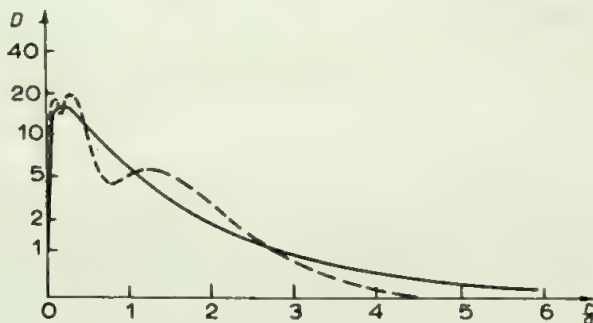


Fig. V.20

The total energy of all electrons in the atom is of the order of magnitude of the mean electrostatic energy of one electron $Ze^2/R \sim Z^{4/3}e^2/a$ multiplied by their total number Z , i.e. is of the order of $e^2Z^{7/3}a$. These mean values, as well as all quantities referring to the properties of the inner regions of atoms, for example the structure of X-ray levels, are in good agreement with experimental data.

On the contrary, quantities which depend on the properties of the peripheral electrons, for example the ionization potentials of the atoms, cannot be determined satisfactorily by the Thomas–Fermi method. At the periphery of the atom the electron density is insufficiently large for the electrons to be considered as a degenerate electron gas.

The main merit of the Thomas–Fermi method is its simplicity. As an example we present an important result which also follows from calculations by the Hartree–Fock method, but which in that case requires very cumbersome calculations. The question is that of finding those values of the atomic number Z for which states with a given value of the orbital angular momentum begin to be occupied. If the electron moves with angular momentum l

in the self-consistent field $\varphi(r)$, then its effective potential energy can be given by formula (35.18). In the quasi-classical approximation $l(l+1)$ can be replaced by $(l+\frac{1}{2})^2$. We then have

$$U_{\text{eff}} = -|e|\varphi(r) + \frac{\hbar^2}{2m} \frac{(l+\frac{1}{2})^2}{r^2},$$

where $\varphi(r)$ is the potential found from the Thomas–Fermi equation. Since the total energy E is always negative, the total potential energy must be essentially negative, $U_{\text{eff}} < 0$, or

$$|e|\varphi(r)r^2 > \frac{\hbar^2}{2m} (l+\frac{1}{2})^2. \quad (71.10)$$

Passing to the dimensionless quantities χ and x , we have instead of (71.10)

$$Z^{\frac{2}{3}} \frac{x\chi(x)}{(\frac{3}{4}\pi)^{\frac{1}{3}}} > (l+\frac{1}{2})^2 \quad (71.11)$$

From fig. V.19 it is seen that the quantity $x\chi(x)$ is limited and has a gently sloping maximum. For large x the potential $\chi(x)$ decreases more rapidly than x^{-1} ; for $x \rightarrow 0$, $x\chi(x)$ is also equal to zero. Hence inequality (71.11) for given l is fulfilled only for a sufficiently large value of Z . This means that the curve U_{eff} lies entirely above the abscissa for a sufficiently small Z and goes below the axis for a sufficiently large Z . There cannot be states with $U_{\text{eff}} > 0$. Hence the limit of realizable states is determined by the condition of the curve U_{eff} being tangent to the abscissa, i.e. by the fulfillment of the conditions

$$U_{\text{eff}} = 0, \quad \frac{dU_{\text{eff}}}{dr} = 0, \quad (71.12)$$

or

$$Z_{\text{crit}}^{\frac{2}{3}} x\chi(x) = \left(\frac{4}{3\pi}\right)^{\frac{2}{3}} (l+\frac{1}{2})^2,$$

$$Z_{\text{crit}}^{\frac{2}{3}} [x^2\chi'(x) - x\chi(x)] = -2 \left(\frac{4}{3\pi}\right)^{\frac{2}{3}} (l+\frac{1}{2})^2.$$

To each value of l there corresponds a certain critical value of nuclear charge Z_{crit} for which the conditions (71.12) are fulfilled.

It is easy to eliminate χ' and χ from these equations, after which we find the relation between l and Z_{crit}

$$Z_{\text{crit}} = 0.155 (2l+1)^3. \quad (71.13)$$

Setting $l = 1, 2, 3, \dots$ in this formula and rounding off the result to the closest integer, we find the values Z_{crit} for which the states with the above angular momenta begin to be occupied. These values are respectively

$$Z_{\text{crit}} = 5, 21, 58, 124. \quad (71.14)$$

In §73 it will be shown that this result is of great importance for understanding the properties of complex atoms.

Another important application of the Thomas–Fermi method is the study of the properties of positive ions. In this case it is to be expected that because of the predominance of the nuclear charge the electron shell will be compressed and the electron density will decrease so rapidly that one can introduce a finite radius of the electron shell, R^* . Outside the ion, for $r > R^*$, an electric field with the potential

$$\varphi = \frac{|e|Z(1-\sigma)}{r}, \quad r > R^*,$$

must exist, where the quantity $\sigma = |\text{charge of the shell}|/\text{charge of the nucleus}$ is called the degree of ionization.

For $r = R^*$ the potential is equal to

$$\varphi_0 = \frac{Z|e|(1-\sigma)}{R^*}.$$

Correspondingly, the energy of an electron at the surface of the ion is equal to $e\varphi_0$.

The condition that the electron be bound in the ion assumes the form

$$E = \frac{p^2}{2m} + e\varphi \leq e\varphi_0$$

instead of $E \leq 0$ for the neutral atom. Correspondingly, the maximum momentum p_{max} is equal to

$$p_{\text{max}} = [2me(\varphi_0 - \varphi)]^{\frac{1}{2}}$$

and eq. (71.5) for the ion assumes the form

$$\frac{1}{r} \frac{d^2(r\varphi)}{dr^2} = - \frac{4e(2me)^{\frac{3}{2}} (\varphi_0 - \varphi)^{\frac{3}{2}}}{3\pi\hbar^3}$$

Its integration taking into account the boundary condition at the surface of the ion $\varphi = \varphi_0$ and condition (71.6) may be carried out numerically, as was

done for the atom. The curve $\chi(x)$ for the rubidium ion is shown in fig. V.19 by the solid line. The curve $\chi(x)$ intersects the abscissa at the point $x^* = R^*/d$, where R^* is determined by the condition

$$4\pi \int_0^{R^*} \rho r^2 dr = -Ze\sigma.$$

Fig. V.21 shows the radial electron density distribution for the ion Rb^+ calculated by the Thomas–Fermi method and according to Hartree (dotted line). We see that in the peripheral part of the ion the agreement between the curves is better than for the atom.

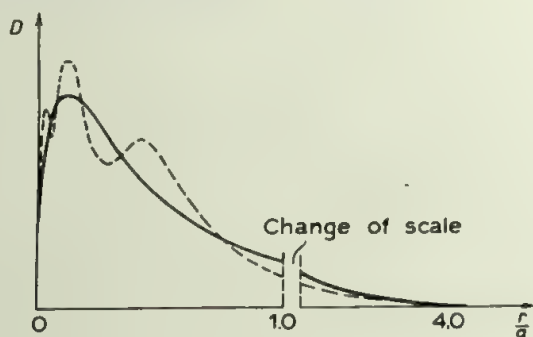


Fig. V.21

The limits of applicability of the Thomas–Fermi method are closely related to those of the quasi-classical approximation.

Upon substituting the expressions

$$U = e\varphi = \frac{Ze^2}{r}, \quad p \approx p_{\max} = (2me\varphi)^{\frac{1}{2}}$$

into formula (39.23) the criterion of applicability of the Thomas–Fermi method is given by the condition

$$r \gg \frac{\hbar^2}{Ze^2m} \sim \frac{a}{Z}.$$

At large distances $r \sim a$ the quasiclassical approximation is invalid. Thus the Thomas–Fermi method is useful for r in the interval

$$a/Z \ll r \ll a. \quad (71.15)$$

§ 72. The quantum numbers characterizing the states of electrons in atoms

We now turn to a discussion of the properties of many-electron atoms. It is obvious that in such an atom (a system consisting of a nucleus and several electrons) the laws of conservation of total energy, total angular momentum, and the component of the angular momentum along an arbitrary axis must be fulfilled. By analogy with the theory of the hydrogen atom one can introduce quantum numbers defining the values of the conserved quantities. At first sight it seems that the quantum numbers must characterize the system as a whole, since, generally speaking, neither the energy nor the angular momentum of an individual electron is conserved. However, the self-consistent field method allows one to consider electrons as independent particles (the wave function of the system is the product of the wave functions of individual particles) in an external field. Each of the electrons moves in the self-consistent spherically symmetric field of the nucleus and of other electrons. Since for motion in a spherically symmetric field the energy, the angular momentum and a component of it along an arbitrary axis are conserved, then not only the atom as a whole but also an individual electron can be characterized by quantum numbers n , l , m . The self-consistent field of the atom is not a Coulomb field, hence the energy levels will depend on l as well as on n . The energy of the electron does not depend on the orientation of the angular momentum in space and, consequently, cannot depend on the quantum number m .

Thus we see that in order to characterize the state of an atom it is necessary to indicate the state of each atomic electron. States with angular momentum $l = 0, 1, 2, 3, \dots$ are denoted respectively by s, p, d, f and so on. The principal quantum number is indicated in the form of a number standing in front of the letter. For example, the notation 5f indicates that in the given state the electron is characterized by the quantum number $n = 5$ and has orbital angular momentum $l = 3$. If several electrons are in a state with the same numbers n and l , then for simplicity their number is indicated in the form of superscript. For example, the normal state of nitrogen is characterized by $1s^2 2s^2 2p^3$. This means that two electrons have the quantum numbers $n = 1, l = 0$; two other electrons are in the state $n = 2, l = 0$ and, finally, three electrons are in the 2p-state. However, this information is insufficient for the complete description of the state of the atom, because it does not tell us how the orbital and spin angular momenta of the individual electrons are combined and what the total angular momentum of the atom is.

We have already mentioned that the total angular momentum of the atom is conserved in time and that because of this it can characterize stationary

states. Furthermore, it can be assumed (neglecting the weak interaction) that the total spin and the total orbital angular momentum of the system are separately conserved. It is just these three quantities which are chosen to characterize the system as a whole. The symbols for the states of atoms with different total angular momenta L are introduced by analogy with the symbols for individual electrons. Namely, for $L = 0, L = 1, L = 2, L = 3$ and so on the states are called respectively S-state, P-state, D-state, F-state and so on. The value of the total angular momentum J is indicated in the form of a subscript on the right of the symbol of the orbital angular momentum. For example, $P_{3/2}$ means that the atom is in a state with orbital angular momentum $L = 1$ and total angular momentum $J = 3/2$. Usually a quantity equal to $2S+1$, where S is the total spin of the atom, is also indicated. The value of $2S+1$ is indicated in the form of a superscript on the left of L . The quantity $2S+1$ for $L \geq S$ shows the number of close levels of the atom constituting its fine structure. Indeed, from the rule of addition of angular momenta it follows that if $L \geq S$, then only $2S+1$ different states may arise when the orbital angular momentum and the spin angular momentum are combined to obtain the total angular momentum of the system.

It turns out that these $2S+1$ states have close but different energies. In other words, $2S+1$ levels form a multiplet. The difference in the energies of the components of the multiplet is associated with the so-called spin-orbit interaction. This is the interaction dependent on the mutual orientation of the orbital and spin angular momentum vectors.

In §118 it will be shown that the relativistic equation for the electron (the Dirac equation) enables one to calculate the spin-orbit interaction.

If, however, one allows for the very fact of the existence of this interaction, i.e. if one assumes that the orientations of the orbital and spin angular momentum vectors are not independent, then the law of interaction can be established from very general considerations. The spin-orbit interaction operator must be a scalar made up of the vectors \mathbf{L} and \mathbf{S} . The only scalar combination is the quantity $\mathbf{L} \cdot \mathbf{S}$. For the mean energy of the spin-orbit interaction we then obtain

$$E_{S-L} = A \overline{\mathbf{L} \cdot \mathbf{S}}$$

where the coefficient A can be either positive or negative. The mean value of $\mathbf{L} \cdot \mathbf{S}$ will be calculated in §74. Formula (74.4) gives

$$E_{S-L} = A' [J(J+1) - L(L+1) - S(S+1)]$$

For different levels belonging to a given multiplet the quantities L and S do not change. Hence for the multiplet splitting we obtain

$$\Delta E_{S-L} = A'J(J+1).$$

The spacing between the neighbouring components of a multiplet is

$$\Delta E = E_{S-L}^{(J)} - E_{S-L}^{(J-1)} = A'J.$$

If $A' > 0$, then the lowest level in the multiplet is the level with the lowest possible value of J (i.e. $J = |L - S|$). These are the so-called normal multiplets. This case is realized for those atoms whose open shell is more than half filled. Otherwise it turns out that $A' < 0$. These are the so-called inverted multiplets, in which the lowest level has the largest total angular momentum J (i.e. $J = L + S$).

The absolute value of multiplet splitting is proportional to Z^2 and rapidly increases in going to heavy atoms.

It is often important to know the total number of possible states of an atom when the quantum numbers n and l of each electron are given. For this purpose it is convenient to use the concept of equivalent electrons which was first introduced by Pauli.

Equivalent electrons are those electrons which have the same quantum numbers n and l .

Where the electrons are not equivalent the calculation of the possible terms is extremely simple.

Let us consider as an example two electrons in a state with $n = 3$, $l = 2$ and $n = 2$, $l = 1$. On the basis of the rule of addition of angular momenta the orbital angular momentum of this system can take on the values $L = 1, 2, 3$, while the total spin of the system can assume two values $S = 0, 1$. Thus we have the terms 1P , 3P , 1D , 3D , 1F , 3F . However for equivalent electrons one has to take into account the Pauli principle in calculating possible terms, and this makes the calculations somewhat more complicated. We consider first the following simple example. Let two electrons be in a state $n_1 = n_2$, and $l_1 = 0$, $l_2 = 0$. In this case the components of the angular momenta in an arbitrary direction are also equal to zero, i.e. $m_1 = 0$, $m_2 = 0$. To satisfy the Pauli principle, s_{z1} and s_{z2} must have opposite signs. Consequently, we may have, for example, $s_{z1} = \frac{1}{2}$, $s_{z2} = -\frac{1}{2}$. But in correspondence with the principle of identity $s_{z1} = -\frac{1}{2}$, $s_{z2} = \frac{1}{2}$ also represents the same state.

The states $s_{z1} = \frac{1}{2}$ and $s_{z2} = \frac{1}{2}$ are forbidden by the Pauli principle. Hence only the term 1S can be realized. The term 3S is forbidden. This calculation shows that He, as well as Be, Mg and Ca and analogous elements cannot have a triplet ground level. We note here that historically Pauli arrived at the exclusion principle by investigating atomic spectra. The exclusion principle

was discovered as a result of the necessity to account for the absence of certain terms.

For what follows we shall need one more example. Let a system of two electrons have the quantum numbers $n_1 = n_2$; $l_1 = l_2 = 1$. In this case each electron can be in the following states:

- | | | |
|----------------------------------|----------------------------------|-----------------------------------|
| 1) $m = 1, s_z = \frac{1}{2}$; | 2) $m = 0, s_z = \frac{1}{2}$; | 3) $m = -1, s_z = \frac{1}{2}$; |
| 4) $m = 1, s_z = -\frac{1}{2}$; | 5) $m = 0, s_z = -\frac{1}{2}$; | 6) $m = -1, s_z = -\frac{1}{2}$. |

In calculating the possible states of the whole system one can combine only different states of individual electrons. This must be done so as not to violate the Pauli principle. According to the rule of addition of angular momenta we obtain the following possible states with $M_z = m_1 + m_2$; $S_z = s_{z1} + s_{z2}$:

- | | | |
|-------------------------|-------------------------|-------------------------|
| 1) $M_z = 2, S_z = 0$; | 2) $M_z = 1, S_z = 1$; | 3) $M_z = 1, S_z = 0$; |
| 4) $M_z = 0, S_z = 0$; | 5) $M_z = 0, S_z = 0$; | 6) $M_z = 0, S_z = 1$; |
| 7) $M_z = 1, S_z = 0$; | 8) $M_z = 0, S_z = 0$. | |

We have not written down analogous states having negative values of the component M_z .

In analyzing the results one has to begin with the state with the largest component M_z . In the case given we have the state with $M_z = 2, S_z = 0$. Hence we conclude that there must be a 1D term (to which there correspond also states with $M_z = 1, S_z = 0$; $M_z = 0, S_z = 0$).

After eliminating the states numbered 1), 3) and 4) from the table we again choose the state with the largest component M_z . In this case $M_z = 1, S_z = 1$. The term 3P corresponds to this state and also to the states denoted by 2), 6), 3) and 4). Finally, only the state $M_z = 0, S_z = 0$, corresponding to the term 1S , remains in the table.

Thus we see that a system of two equivalent electrons with $l_1 = 1, l_2 = 1$ can be in the states $^1D, ^3P$ and 1S . In computing states in more complex cases one has to proceed in an analogous way.

We now discuss some general regularities in the ordering of the energy levels of an atom. If the electrons are in states with definite numbers n and l (in such cases one says that the electron configuration is given), then to such a distribution there may correspond several different energy levels differing in the total orbital angular momentum as well as in the total spin of the system.

Taking into account the multiplet splitting, the state of the atom turns out to depend on the quantities J, L and S .

The state of the atom as a whole, represented by the atomic term, is determined by these quantities. The symbol for the term is $^{2S+1}L_J$. For example, the normal term of the nitrogen atom ($L = 0$, $S = \frac{3}{2}$, $J = \frac{3}{2}$) is written as $^4S_{3/2}$.

The ordering of terms of different multiplicity was obtained from calculations carried out by the Hartree–Fock method (although historically it was established much earlier by Hund).

It turns out that of all the terms of a given configuration the lowest energy is possessed by the term with the largest value of the total spin S .

For a given S the term with the largest value of L has the lowest energy.

In the example just given of the terms 1D , 1S and 3P the order of the terms in increasing energy will be 3P , 1D and 1S . As to the ordering of levels within a given multiplet, there are two cases. The first of these, called the multiplet with normal structure, is characterized by an increase in the energy of the levels with increasing total angular momentum L . In the second case, where the energy of the levels decreases with increasing L , the structure of the multiplet is said to be inverted. It turns out that if the number of equivalent electrons in the atom or ion is lower than the total number of electrons, then the multiplets have a normal structure. In atoms or ions in which the number of equivalent electrons is larger than or equal to one half of the total number of electrons the multiplets are inverted. For example, in the oxygen atom of the eight electrons four are in the $2p$ -state (structure $2p^4$) and are equivalent. Hence in the case of the oxygen atom the multiplets are inverted. In the oxygen ion O^{2-} there are 2 electrons in the $2p$ -state (configuration $2p^2$), and the multiplets have a normal structure.

We now introduce the concept of an atomic shell. This is the set of all the electron states with the same values of the quantum numbers n and l . If all states with the quantum numbers n and l are occupied, then the corresponding shell is closed. It is known that for given n and l there are in all $2l+1$ different states differing in the quantum number m . If the spin is also taken into account, then the total number of electrons necessary to fill the shell will be equal to $2(2l+1)$. If the shell is closed, then the total spin of the system, as well as the components of the orbital angular momentum, must be equal to zero. In this case $S = 0$, $L = 0$, $J = 0$. This can be shown by taking into account the Pauli principle and recalling that in a completely filled shell all possible states with positive as well as negative projections of the angular momentum onto the z -axis are occupied. The term 1S_0 corresponds to a closed shell.

We stress that our preceding considerations were based on the assumption that the orbital angular momenta of the electrons were combined into a total orbital angular momentum of the system, and that the spin angular momenta

were combined into a total spin angular momentum of the system. Such an assumption corresponds to the statement that the interaction between the spin and orbital motions of electrons is much weaker than the interaction between the spins. Then one can speak of approximate conservation of the total orbital angular momentum and of the total spin of the system. This type of interaction is called normal or Russell—Saunders coupling. On the basis of the assumption of normal coupling it turns out to be possible to systematize the lowest energy levels of most atoms. Deviations from normal coupling are observed in the seventh and eighth groups of the periodic system of the elements.

In principle another limiting form of coupling, usually called *jj* coupling, is also possible. In *jj* coupling the orbital angular momentum and the spin angular momentum of each electron are added to give the total angular momentum j of that electron (in this case the orbital angular momentum of an individual electron is not conserved). In their turn the total angular momenta of the individual electrons are added together to give the total angular momentum of the atom, J . Such coupling is not encountered in atoms in its pure form.

Let us consider some examples of different modifications of the basic forms of coupling in atoms. If an electron is in a highly excited state and, consequently, is sufficiently far from the nucleus and other electrons of the atom, then the behaviour of this electron can be considered to be independent of the rest of the atom. In this case the total angular momentum of the individual electron can be considered to be conserved independently of the total angular momentum of the rest of the atom.

Consider another example. For atoms with a large charge Z the inner electrons interact strongly with the nuclear charge and relatively weakly with the outer atomic electrons. Hence one can assume approximately that the inner electrons do not interact with the other electrons (the total angular momentum of such electrons is conserved). In this case one can speak of *jj* coupling. We note that such electrons must be characterized not by the quantum numbers n and l but by the quantum numbers n and j .

For use in what follows we shall now show that the electric dipole moment of an atom

$$\mathbf{d} = \sum_{i=1}^N \int e_i \mathbf{r}_i |\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i, \dots, \mathbf{r}_N)|^2 dV_1 dV_2 \dots dV_N, \quad (72.1)$$

in a stationary state with definite parity is equal to zero. Indeed, since the

parity operator commutes with the Hamiltonian, the wave function ψ is an eigenfunction of the parity operator. In other words, it satisfies the relation

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = -\psi(-\mathbf{r}_1, -\mathbf{r}_2, \dots, -\mathbf{r}_n)$$

or

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = \psi(-\mathbf{r}_1, -\mathbf{r}_2, \dots, -\mathbf{r}_n).$$

In both cases the function $|\psi|^2$ is an even function. It is now obvious that the integrand in (72.1) is odd and, consequently, the dipole moment of an atom is equal to zero.

§73. The periodic system of the elements

In its time the theoretical construction of the Mendeleyev periodic system, carried out by Bohr in 1922, was one of the most effective results obtained by means of the quantum theory.

The construction of the periodic system of the elements is based on three assumptions:

- (1) The structure of atoms is determined by the atomic number Z (the charge of the nucleus).
- (2) As the atomic number and the number of electrons in the atom increase the electrons fill the states with the lowest possible energy.
- (3) The occupation of energy states is limited by the Pauli principle.

In §72 we have defined the term atomic shell. A closed shell contains $2(2l+1)$ electrons. The energy of an atom depends only on the quantum numbers n and l . Thus all $2(2l+1)$ electrons in a shell have the same energy (if we do not take into account the weak spin-orbit interaction).

The set of sub-shells with fixed principal quantum number n is called a shell. The number of electrons filling a shell is equal to

$$2 \sum_{l=0}^{n-1} (2l+1) = 2n^2.$$

Each shell is denoted by letters taken from the classification adopted in X-ray spectroscopy. That is, as follows:

n	1	2	3	4	5
Symbol of the shell	K	L	M	N	O
Possible number of electrons in the shell	2	8	18	32	50

In contrast to the hydrogen atom, in other atomic systems, the energy of states is defined by both the principal quantum number n and the orbital number l . The dependence of the energy on n is, generally speaking, stronger than the dependence on l . This means that for all values of l states with a given value of n lie lower than states with the quantum number $n + 1$. The sequence of energy states is

$$1s, 2s, 2p, 3s, 3p, \dots$$

However, in going to d-states and particularly to f-states the situation changes. For large values of the angular momentum the dependence of the energy on the orbital quantum number l turns out to be most important.

The effective potential energy of the electrons arises from the Coulomb field of the nucleus screened by electrons, and the centrifugal force. The screened potential of the nucleus decreases at large distances substantially more slowly than the Coulomb potential and still more slowly than the centrifugal potential.

Comparison between the total effective energy of electrons with small angular momenta (s- and p-states) and large angular momenta (d- and f-states) shows that there is an essential difference between them. Namely, the curve U_{eff} for $l = 2$ and 3 lies higher than for states with $l = 0$ and 1.

Because of this the minimum of the energy for d- and f-states lies closer to the nucleus than for s- and p-states. This means that on the average d-electrons and particularly f-electrons move closer to the nucleus, in deeper parts of the shell, than s-electrons and p-electrons; d-electrons and f-electrons are often said to be penetrating. This general property of states with large angular momenta leads to the fact that 3d-electrons on the average move closer to the nucleus than 4s-electrons.

On the other hand, the energy in the screened field increases with increasing angular momentum. Experiment and calculations by the Hartree-Fock method show that the energy of the 4s-state lies below the energy of the 3d-state. Hence the order of filling states lying above 3p turns out to be as follows:

$$4s, 3d, 4p, 5s, 4d, 5p, 6s, 4f, 5d, 6p, 7s, 6d, 5f.$$

Elements in which the 3d-shell and particularly the 4f-shell and the 5f-shell

are partly filled possess special properties. Since the motion in a non-Coulomb field penetrates closer to the nucleus for states with large angular momenta, the addition of electrons to the d-shell and particularly to the f-shell does not change those properties of atoms which depend on the peripheral electrons.

Let us analyze in more detail the order in which states are occupied. This will allow us to find out which properties of atoms should display a periodic trend and which a monotonic trend with increasing atomic number Z .

The first element of the periodic system is hydrogen. Its normal term is $2S_{1/2}$.

In the next element (helium) the K-shell having two electrons is filled. It is easy to find that, in accordance with the rules discussed in the preceding section, the normal term of helium is $1S_0$.

The building of the L-shell begins in the next element (lithium). The third electron of lithium goes into a 2s-state.

In calculating the normal term one need not take into account the electrons of the filled shell: their spin orbital and total angular momenta are equal to zero.

The normal term of lithium is defined by the single electron of the L-shell. Lithium is in the state $2S_{1/2}$.

In beryllium the fourth electron fills the 2s-shell. The normal term of beryllium is $1S_0$, as for helium.

The fifth electron of boron goes into the 2p-state. Thus the boron atom has the following distribution of electrons: $1s^2 2s^2 2p$. Since the 1s and 2s shells are filled, the normal term of the boron atom is easily found. It is $2P_{1/2}$.

In the case of carbon, six electrons are distributed as follows: $1s^2 2s^2 2p^2$. In order to determine the normal term of carbon, we turn to the example discussed in § 72 of the determination of terms for two equivalent p-electrons. Making use of the Hund rule, we see that the normal term of carbon is $3P$. In this case the atom contains less than one half of all possible equivalent p-electrons. Hence the multiplet structure of the lower level corresponds to a minimum J , in the given case $J = 0$. Thus, finally, for the normal term we have the symbol $3P_0$.

The order of further filling the terms of normal states is shown in table 2. We see that in neon the L-shell is complete and that, like helium, it has no electrons in unfilled shells and sub-shells.

It is natural to identify filled shells with the periods of the Mendeleyev system of the elements. Each period begins to be filled by one electron in an s-state and is completed when a filled shell is formed.

The first period of the periodic system of the elements contains elements in which the K-shell is filled. It comprises two elements ($n=1, l=0$). The second

Table 2
The distribution of electrons in the periodic system of elements

Element		K	L		M			N		Normal term	Ionization potential (eV)
		1s	2s	2p	3s	3p	3d	4s	4p		
H	1	1	—	—	—	—	—	—	—	$^2S_{1/2}$	13.595
He	2	2	—	—	—	—	—	—	—	1S_0	24.58
Li	3	2	1	—	—	—	—	—	—	$^2S_{1/2}$	5.39
Be	4	2	2	—	—	—	—	—	—	1S_0	9.32
B	5	2	2	1	—	—	—	—	—	$^2P_{1/2}$	8.30
C	6	2	2	2	—	—	—	—	—	3P_0	11.26
N	7	2	2	3	—	—	—	—	—	$^4S_{3/2}$	14.53
O	8	2	2	4	—	—	—	—	—	3P_2	13.61
F	9	2	2	5	—	—	—	—	—	$^2P_{3/2}$	17.42
Ne	10	2	2	6	—	—	—	—	—	1S_0	21.56
Na	11	Neon configuration			1	—	—	—	—	$^2S_{1/2}$	5.14
Mg	12				2	—	—	—	—	1S_0	7.64
Al	13				2	1	—	—	—	$^2P_{1/2}$	5.98
Si	14				2	2	—	—	—	3P_0	8.15
P	15				2	3	—	—	—	$^4S_{3/2}$	10.48
S	16				2	4	—	—	—	3P_2	10.36
Cl	17				2	5	—	—	—	$^2P_{3/2}$	13.01
Ar	18				2	6	—	—	—	1S_0	15.76
K	19	Argon configuration					—	1	—	$^2S_{1/2}$	4.34
Ca	20						—	2	—	1S_0	6.11
Sc	21						1	2	—	$^2D_{3/2}$	6.54
Ti	22						2	2	—	3F_2	6.82
V	23						3	2	—	$^4F_{3/2}$	6.74
Cr	24						5	1	—	7S_3	6.76
Mn	25						5	2	—	$^6S_{5/2}$	7.43
Fe	26						6	2	—	5D_4	7.90
Co	27	Argon configuration					7	2	—	$^4F_{9/2}$	7.86
Ni	28						8	2	—	3F_4	7.63
Cu	29						10	1	—	$^2S_{1/2}$	7.72
Zn	30						10	2	—	1S_0	9.39
Ga	31						10	2	1	$^2P_{1/2}$	6.00
Ge	32						10	2	2	3P_0	7.88
As	33						10	2	3	$^4S_{3/2}$	9.81
Se	34						10	2	4	3P_2	9.75
Br	35						10	2	5	$^2P_{3/2}$	11.84
Kr	36						10	2	6	1S_0	14.00

Table 2 (continued)

Element	Configuration of inner shells	N		O			P	Normal term	Ionization potential (eV)
		4d	4f	5s	5p	5d	6s		
Rb 37	Krypton configuration	—	—	1	—	—	—	$^2S_{1/2}$	4.19
Sr 38		—	—	2	—	—	—	1S_0	5.69
Y 39		1	—	2	—	—	—	$^2D_{3/2}$	6.38
Zr 40		2	—	2	—	—	—	3F_2	6.84
Nb 41		4	—	1	—	—	—	$^6D_{1/2}$	6.88
Mo 42		5	—	1	—	—	—	7S_3	7.10
Tc 43		5	—	2	—	—	—	$^6S_{5/2}$	7.28
Ru 44		7	—	1	—	—	—	5F_5	7.36
Rh 45	8	—	1	—	—	—	$^4F_{9/2}$	7.46	
Pd 46	10	—	—	—	—	—	1S_0	8.33	
Ag 47	Paladium configuration	—	—	1	—	—	—	$^2S_{1/2}$	7.57
Cd 48		—	—	2	—	—	—	1S_0	8.99
In 49		—	—	2	1	—	—	$^2P_{1/2}$	5.78
Sn 50		—	—	2	2	—	—	3P_0	7.34
Sb 51		—	—	2	3	—	—	$^4S_{3/2}$	8.64
Te 52		—	—	2	4	—	—	3P_2	9.01
I 53		—	—	2	5	—	—	$^2P_{3/2}$	10.45
Xe 54		—	—	2	6	—	—	1S_0	12.13
Cs 55	The sub-shells from 1s to 4d contain 46 electrons	—	—	The 5s and 5p sub-shells together contain 8 electrons	—	1	$^2S_{1/2}$	3.89	
Ba 56		—	—		2	1S_0	5.21		
La 57		—	—		1	2	$^2D_{3/2}$	5.61	
Ce 58		2	—		2	3H_4	6.91		
Pr 59		3	—		2	$^4I_{9/2}$	5.76		
Nd 60		4	—		2	5I_4	6.31		
Pm 61		5	—		2	$^6H_{5/2}$	6.30		
Sm 62		6	—		2	7F_0	5.10		
Eu 63		7	—		2	$^8S_{7/2}$	5.67		
Gd 64		7	1		2	9D_2	11.40		
Tb 65		8	1		2	$^8H_{17/2}$	6.74		
Dy 66		10	—		2	5I_8	6.82		
Ho 67		11	—		2	$^4I_{15/2}$	6.90		
Er 68		12	—		2	3H_6	6.90		
Tm 69		13	—		2	$^2F_{7/2}$	6.90		
Yb 70		14	—		2	1S_0	6.20		
Lu 71		14	1		2	$^2D_{3/2}$	5.00		

Table 2 (continued)

Element	Configuration of inner shells	O		P			Q	Normal term	Ionization potential (eV)
		5d	5f	6s	6p	6d			
Hf 72	The sub-shells from 1s to 5p contain 68 electrons	2	—	2	—	—	—	3F_2	7.00
Ta 73		3	—	2	—	—	—	$^4F_{3/2}$	7.88
W 74		4	—	2	—	—	—	5D_0	7.98
Re 75		5	—	2	—	—	—	$^6S_{5/2}$	7.87
Os 76		6	—	2	—	—	—	5D_4	8.70
Ir 77		7	—	2	—	—	—	$^4F_{1/2}$	9.00
Pt 78		9	—	1	—	—	—	3D_3	9.00
Au 79	The sub-shells from 1s to 5d contain 78 electrons	—	—	1	—	—	—	$^2S_{1/2}$	9.22
Hg 80		—	—	2	—	—	—	1S_0	10.44
Tl 81		—	—	2	1	—	—	$^2P_{1/2}$	6.11
Pb 82		—	—	2	2	—	—	3P_0	7.42
Bi 83		—	—	2	3	—	—	$^4S_{3/2}$	7.29
Po 84		—	—	2	4	—	—	3P_2	8.43
At 85		—	—	2	5	—	—	$^2P_{3/2}$	9.40
Rn 86		—	—	2	6	—	—	1S_0	10.75
Fr 87	The sub-shells from 1s to 5d contain 78 electrons	—	—	2	6	—	1	$^2S_{1/2}$	4.00
Ra 88		—	—	2	6	—	2	1S_0	5.28
Ac 89		—	—	2	6	1	2	$^2D_{3/2}$	5.5
Th 90		—	—	2	6	2	2	3F_2	5.7
Pa 91		2	—	2	6	1	2	$^4K_{11/2}$	5.7
U 92		3	—	2	6	1	2	5L_6	4.0
Np 93		4	—	2	6	1	2	$^6L_{11/2}$	
Pu 94		6	—	2	6	—	2	7F_0	
Am 95		7	—	2	6	—	2	$^8S_{7/2}$	
Cm 96		7	—	2	6	1	2	9D_2	
Bk 97		8	—	2	6	1	2	$^8H_{17/2}$	
Cf 98		10	—	2	6	—	2	6I_8	
Es 99		11	—	2	6	—	2	$^4I_{15/2}$	
Fm 100		12	—	2	6	—	2	3H_6	
Md 101		13	—	2	6	—	2	$^2F_{7/2}$	
No 102		14	—	2	6	—	2	1S_0	
Lw 103		14	—	2	6	1	2	$^2D_{1/2}$	
Ku 104		14	—	2	6	2	2	3F_2	

period contains elements with the L-shell filled. It comprises 8 elements ($n=2$, $l=0,1$) from lithium to neon. The 3s-states and 3p-states of the M-shell are filled from sodium to argon. Up to now periods ending with the noble gases He, Ne, Ar have been formed in the Mendeleyev periodic system. In the next period, beginning with potassium and ending with krypton, there is a departure from the simple rules of successive filling. Namely, as we have seen in § 71, electrons with angular momentum $l = 2$, i.e. in the d-state, must begin at the element with $Z = 21$. Hence in Sc the additional twenty-first electron does not go into a 4p-state but into a 3d-state, which proceeds to fill from Sc to Ni. It is interesting to note that in Cr the tendency to fill a 3d-state is so strong that one of the 4s-electrons goes over into a 3d-state. The filling of 4s-states and 4p-states of the fourth period of the system of elements, which ends with krypton, begins again after Ni.

The further simple building of the N-shell, i.e. the fifth period containing 18 elements up to Xe, proceeds after krypton. In the sixth period, containing 32 elements, the filling of 6s-, 4f- and 6p-states proceeds. Here again there is a more complex order of filling. In Ce ($Z=58$) the electrons begin to fill the 4f-states. We note that according to a calculation in the Thomas—Fermi approximation the electrons with $l = 3$ should appear beginning with the element $Z = 55$.

The building of the seventh period for elements existing in nature remains incomplete. The filling of the deep 5f-state begins with Pa. Up to now there are artificially produced elements from Np ($Z = 93$) to kourchatovium (Ku, $Z=104$). In all these elements, called the actinides, the filling of the 5f-states proceeds.

As the atomic number Z increases all the properties of atoms determined by the inner electrons display monotonic changes. As an example we can cite the characteristic X-ray spectra. Characteristic X-rays arise when a vacancy in one of the inner shells is filled. The X-ray spectrum evidently has a character similar to that of hydrogen, but with the Rydberg constant multiplied by Z^2 . The frequency of emission lines increases in proportion to Z^2 (the Moseley law).

On the other hand, all the properties of atoms determined by the peripheral electrons have a periodic behaviour with increasing Z . For example, the ionization potential of the atom (see table 2) is one such property. It has its smallest value for the first element and reaches its largest value for the last element of a period.

Another property displaying periodicity is the atomic volume. The largest volume is possessed by the alkali metal atoms, in which there is one electron outside a filled shell.

From the point of view of chemistry the most important characteristic of an atom is its valence. Chemical binding only involves unpaired electrons. Electrons which are in filled states and have a total spin equal to zero do not take part in the chemical interaction.

Hence it follows that the chemical interaction and its qualitative characteristic (the valence) are determined exclusively by the number of unpaired electrons which are in unfilled states. The numerical value of the valence of an atom in a given state is equal to $r = 2S$, where S is the spin of the atom in that state.

We particularly stress that the valence of an atom is related to its state because the atom, when making a transition from one state into another, may change its valence.

In §79 we shall discuss this problem in somewhat more detail. Here we shall confine ourselves to the following remark: if the first excited term lies close to the normal term, then the atom may be involved in chemical binding in the excited state. From the above it follows that the elements of the first group of the periodic system, with which all 7 periods begin, in the $^2S_{1/2}$ state have the valence $r = 1$.

The elements of the second group which are in the 1S_0 state have zero valence. They would be chemically inert if the excited term did not lie close to the normal state. In the excited state the two outer electrons have the configuration s and p , so that the atom has a total spin 1 and a valence $r = 2$.

In the atoms of the third group three electrons are outside filled shells. Their configuration s^2p corresponds to a total spin $S = \frac{1}{2}$ and to a valence one. However, those of these atoms which have a small excitation energy may make a transition into the state with the configuration sp^2 and spin $S = \frac{3}{2}$. In the excited state their valence is $r = 3$. The elements of the first three groups are regarded chemically as metals. From the chemical point of view metals are characterized by the ability to lose electrons when forming ionic chemical compounds.

The elements of the fourth group enter into a chemical bond in the normal and excited states with the configurations s^2p^2 and sp^3 . The corresponding values of the spin and valence are equal respectively to $S = 1$, $r = 2$ and $S = 2$, $r = 4$.

The excited state corresponds to the configuration s^2p^2s and to the transition of the fifth electron into the s -state of the next shell (i.e. to a transition with an increase in the principal quantum number by one). The spin and valence in the excited state are equal respectively to $S = \frac{5}{2}$ and $r = 5$.

In the sixth group the atoms in the normal state have the configuration s^2p^4 with spin $S = 1$. Their valence is $r = 2$. When excited, one of the elec-

trons makes a transition from the p-state to the s-state of the next shell. In this excited state $r = 4$.

In addition to this type of excitation, excitations with a transition of two electrons into the next shell, one from the s-state and the second from the p-state, are often brought about. In this excited state the atom has the valence $r = 6$.

In the atoms of the seventh group the normal configuration is s^2p^5 , the spin is $S = \frac{1}{2}$ and the valence is $r = 1$. However, transitions of one, two and three electrons into the next layer are possible. Hence also the valences $r = 3$, $r = 5$ and $r = 7$ are realized.

The elements of the fourth, fifth, sixth and seventh groups which stand at the beginning of the groups are non-metals. In compounds of ionic type they gain electrons (are oxidizers), having the tendency to form a filled state.

The elements of the transition groups — iron, palladium and platinum — as well as the lanthanides (rare earths) and actinides, have special chemical properties.

The completion of deep d-states and f-states takes place in the atoms of the group of iron, the lanthanides and the actinides, d-electrons and f-electrons usually do not take part in valence bonds and the valence of the atoms is determined by the electrons in the outer states. However, this is not a strict law, since in certain cases of chemical compound formation, electrons from inner states make transitions into outer states and contribute to the valence. This is particularly clearly displayed in the case of some actinides. Hence the chemical properties of elements of the groups with special properties are rather complex.

Thus we see that not only is the theoretical substantiation of the distribution of atoms in the periodic system of the elements possible, but further a relatively detailed prediction of their chemical properties can be given.

§74. The Zeeman effect

We have seen in §31 of Part I that the full theory of the Zeeman effect could not be constructed on the basis of classical electrodynamics. After analyzing a vast amount of experimental data, Landé introduced a parameter which quantitatively determines the characteristics of the Zeeman effect. This parameter is called the Landé g -factor.

The quantum theory of the Zeeman effect makes it possible to find the value of the Landé g -factor and the form of the Zeeman splitting without any new assumptions. Let us consider how the positions of the energy levels

of an atom change if the atom is placed in a constant external magnetic field. The wave function ψ for the stationary states of an atom can, as usual, be written in the form

$$\psi(\mathbf{r}_i, t) = \psi(\mathbf{r}_i) e^{-iEt/\hbar} . \quad (74.1)$$

On substituting (74.1) into the Pauli equation (63.11) the latter transforms into the form

$$\left\{ \frac{1}{2m} \sum_i \hat{\mathbf{p}}_i^2 + \frac{|e|\hbar}{2mc} (\hat{\mathbf{L}} + 2\hat{\mathbf{S}}) \cdot \boldsymbol{\mathcal{A}} + \frac{e^2}{8mc^2} \sum_i [\boldsymbol{\mathcal{A}} \times \mathbf{r}_i]^2 + U \right\} \psi = E\psi , \quad (74.2)$$

where U takes into account the interaction of the electrons with each other and with the nuclei. We assume that the external magnetic field strength is sufficiently small that the term containing the square of the field can be dropped in (74.2).

We introduce the quantity \hat{H}' given by

$$\hat{H}' = \frac{|e|\hbar}{2mc} (\hat{\mathbf{L}} + 2\hat{\mathbf{S}}) \cdot \boldsymbol{\mathcal{A}} = \frac{-e\hbar}{2mc} (\hat{\mathbf{J}} + \hat{\mathbf{S}}) \cdot \boldsymbol{\mathcal{A}} , \quad (74.3)$$

where $\hat{\mathbf{J}}$ is the total angular momentum operator, \hat{H}' is the small perturbation acting on the atom.

The Hamiltonian \hat{H} can then be written in the form

$$\hat{H} = \hat{H}_0 + \hat{H}' ; \quad \hat{H}_0 = \frac{1}{2m} \sum_i \hat{\mathbf{p}}_i^2 + U . \quad (74.4)$$

In the unperturbed state the atom is characterized by a definite total angular momentum of the system J and by a definite z -component of the total angular momentum, M_z . Clearly, one has to apply perturbation theory for degenerate states. Indeed, the energy of the unperturbed state does not depend on the value of the total angular momentum component M_z . Since the perturbation operator represents the projection of a certain vector onto the z -axis and is brought to diagonal form simultaneously with the operator of the z -component of the total angular momentum, one needs to calculate only the diagonal elements of the perturbation operator

$$\hat{H}' = \frac{|e|\hbar}{2mc} (\hat{J}_z + \hat{S}_z) . \quad (74.5)$$

The diagonal matrix element is taken with respect to the quantum numbers

of the total angular momentum J and the z -component of the angular momentum $M_z \equiv M$.

The diagonal matrix element of the operator \hat{J}_z is equal to

$$(\hat{J}_z)_{JM;JM} = \hbar M. \quad (74.6)$$

Consequently, we have to determine the expression

$$(\hat{S}_z)_{JM;JM} = \bar{S}_z.$$

The calculation of this matrix element by means of the commutation rules is a good example of the practical use of the matrix method. We note that the value of \bar{S}_z is usually found from obvious but not quite rigorous considerations associated with the precession of the vector \mathbf{S} with respect to the vector \mathbf{J}^* . We note at first that by definition the operator $\hat{\mathbf{J}}$ is equal to $\hat{\mathbf{J}} = \hat{\mathbf{L}} + \hat{\mathbf{S}}$. The operators $\hat{\mathbf{L}}$ and $\hat{\mathbf{S}}$ commute with each other, since they act on different variables.

Knowing the commutation rules for $\hat{L}_x, \hat{L}_y, \hat{L}_z; \hat{S}_x, \hat{S}_y, \hat{S}_z$ we easily find the following relations:

$$\{\hat{J}_x, \hat{S}_x\} = 0, \quad \{\hat{J}_x, \hat{S}_y\} = i\hbar \hat{S}_z, \quad \{\hat{J}_x, \hat{S}_z\} = -i\hbar \hat{S}_y.$$

Other commutation rules can be obtained by cyclic permutation. Then we obtain

$$\begin{aligned} \{\hat{J}_y, \hat{S}_y\} &= 0, & \{\hat{J}_y, \hat{S}_z\} &= i\hbar \hat{S}_x, & \{\hat{J}_y, \hat{S}_x\} &= -i\hbar \hat{S}_z, \\ \{\hat{J}_z, \hat{S}_z\} &= 0, & \{\hat{J}_z, \hat{S}_y\} &= -i\hbar \hat{S}_x, & \{\hat{J}_z, \hat{S}_x\} &= i\hbar \hat{S}_y. \end{aligned} \quad (74.7)$$

From these rules there results the relation

$$(\hat{J}_x + i\hat{J}_y)(\hat{S}_x + i\hat{S}_y) - (\hat{S}_x + i\hat{S}_y)(\hat{J}_x + i\hat{J}_y) = 0. \quad (74.8)$$

We calculate the following matrix element of the right-hand and left-hand sides of this relation:

$$[(\hat{J}_x + i\hat{J}_y)(\hat{S}_x + i\hat{S}_y)]_{J,M+1;J,M-1} = [(\hat{S}_x + i\hat{S}_y)(\hat{J}_x + i\hat{J}_y)]_{J,M+1;J,M-1}.$$

In correspondence with formula (51.14) the matrix element of the operator $\hat{J}_x + i\hat{J}_y$ differs from zero only in the case of the transition $J, M \rightarrow J, M-1$. Hence

$$(\hat{J}_x + i\hat{J}_y)_{JM;JM-1} = \hbar [(J+M)(J-M+1)]^{\frac{1}{2}}. \quad (74.9)$$

Then making use of the rule of multiplication for matrices (45.6), we obtain

* See L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

$$(\hat{J}_x + i\hat{J}_y)_{M+1;M}(\hat{S}_x + i\hat{S}_y)_{M;M-1} - (\hat{S}_x + i\hat{S}_y)_{M+1;M}(\hat{J}_x + i\hat{J}_y)_{M;M-1} = 0. \quad (74.10)$$

In this formulae we have dropped the suffix J . Using relation (74.9) we find

$$\frac{(\hat{S}_x + i\hat{S}_y)_{M+1;M}}{[(J+M+1)(J-M)]^{\frac{1}{2}}} = \frac{(\hat{S}_x + i\hat{S}_y)_{M;M-1}}{[(J+M)(J-M+1)]^{\frac{1}{2}}} \equiv A.$$

Analogously one can obtain

$$\frac{(\hat{S}_x + i\hat{S}_y)_{M+2;M+1}}{[(J+M+2)(J-M-1)]^{\frac{1}{2}}} = A.$$

Hence we see that the quantity A does not depend on M and, consequently, we have

$$(\hat{S}_x + i\hat{S}_y)_{M;M-1} = A[(J+M)(J-M+1)]^{\frac{1}{2}}. \quad (74.11)$$

The matrix elements of the operator \hat{S}_z which we need can be found by making use of the following formula:

$$(\hat{J}_x - i\hat{J}_y)(\hat{S}_x + i\hat{S}_y) - (\hat{S}_x + i\hat{S}_y)(\hat{J}_x - i\hat{J}_y) = -2\hbar\hat{S}_z. \quad (74.12)$$

Calculating the diagonal matrix element of relation (74.12), we obtain as a result

$$\begin{aligned} (\hat{J}_x - i\hat{J}_y)_{M;M+1}(\hat{S}_x + i\hat{S}_y)_{M+1;M} - (\hat{S}_x + i\hat{S}_y)_{M;M-1}(\hat{J}_x - i\hat{J}_y)_{M-1;M} = \\ = -2\hbar(\hat{S}_z)_{M;M}. \end{aligned}$$

Making use of (74.11) and knowing that

$$(\hat{J}_x - i\hat{J}_y)_{M;M+1} = \hbar[(J+M+1)(J-M)]^{\frac{1}{2}},$$

we can easily determine the diagonal element $(\hat{S}_z)_{M;M}$ which is equal to

$$\begin{aligned} -2(\hat{S}_z)_{M;M} &= (J+M+1)(J-M)A - (J+M)(J-M+1)A = -2AM, \\ (\hat{S}_z)_{M;M} &= AM. \end{aligned} \quad (74.13)$$

We now turn to finding the quantity A . From the relation

$$\hat{J}^2 = (\hat{L} + \hat{S})^2 = \hat{L}^2 + 2(\hat{L} \cdot \hat{S}) + \hat{S}^2 = \hat{L}^2 + 2(\hat{J} \cdot \hat{S}) - \hat{S}^2$$

it immediately follows that

$$\hat{J} \cdot \hat{S} = \frac{1}{2}(\hat{J}^2 + \hat{S}^2 - \hat{L}^2).$$

In the case of Russell-Saunders coupling the diagonal matrix element of the scalar product $\hat{J} \cdot \hat{S}$ is equal to

$$(\hat{\mathbf{J}} \cdot \hat{\mathbf{S}})_{JM; JM} = \frac{1}{2} \hbar^2 [J(J+1) - L(L+1) + S(S+1)] . \quad (74.14)$$

On the other hand, the matrix element can be found if the scalar expression $\hat{\mathbf{J}} \cdot \hat{\mathbf{S}}$ is transformed to the form

$$(\hat{\mathbf{J}} \cdot \hat{\mathbf{S}})_J = \frac{1}{2} (\hat{S}_x + i\hat{S}_y)(\hat{J}_x - i\hat{J}_y) + \frac{1}{2} (\hat{S}_x - i\hat{S}_y)(\hat{J}_x + i\hat{J}_y) + \hat{J}_z \hat{S}_z . \quad (74.15)$$

We find the diagonal matrix element of the right-hand and left-hand sides of relation (74.15). We then have

$$\begin{aligned} (\hat{\mathbf{J}} \cdot \hat{\mathbf{S}})_{M; M} &= \frac{1}{2} (\hat{S}_x + i\hat{S}_y)_{M; M-1} (\hat{J}_x - i\hat{J}_y)_{M-1; M} + \\ &+ \frac{1}{2} (\hat{S}_x - i\hat{S}_y)_{M; M+1} (\hat{J}_x + i\hat{J}_y)_{M+1; M} + \hbar M (\hat{S}_z)_{M; M} . \end{aligned} \quad (74.16)$$

We now need to find the matrix element $(\hat{S}_x - i\hat{S}_y)_{M; M+1}$. This is easily done by means of relation (74.11). We note beforehand that the constant A is real. This is seen from formula (74.13), in which all the quantities determining the quantity A are real.

Carrying out the complex conjugation of the right-hand and left-hand sides of the relation

$$(\hat{S}_x + i\hat{S}_y)_{M+1; M} = A [(J+M+1)(J-M)]^{\frac{1}{2}} ,$$

we obtain

$$(\hat{S}_x)_{M+1; M}^* - i(\hat{S}_y)_{M+1; M}^* = A [(J+M+1)(J-M)]^{\frac{1}{2}} .$$

Making use of the hermiticity of the operators, we find

$$\begin{aligned} (\hat{S}_x)_{M+1; M}^* - i(\hat{S}_y)_{M+1; M}^* &= (\hat{S}_x)_{M; M+1} - i(\hat{S}_y)_{M; M+1} = \\ &= (\hat{S}_x - i\hat{S}_y)_{M; M+1} = A [(J+M+1)(J-M)]^{\frac{1}{2}} . \end{aligned} \quad (74.17)$$

Substituting the values of $(\hat{\mathbf{J}} \cdot \hat{\mathbf{S}})_{M; M}$ and $(\hat{S}_x + i\hat{S}_y)_{M; M-1}$ into relation (74.16) in correspondence with formulae (74.14) and (74.11), and also using (74.17) and (74.13), we find the quantity A :

$$A = \hbar \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} . \quad (74.18)$$

Using the value found for A and also formula (74.13), we obtain the diagonal matrix element $(\hat{S}_z)_{JM; JM}$ which has the form

$$(\hat{S}_z)_{JM; JM} = \hbar M \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} . \quad (74.19)$$

The correction to the energy levels of the atom due to the magnetic field \mathcal{H} is given by the expression

$$\Delta E = \frac{|e|\mathcal{H}\hbar M}{2mc} \left(1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \right) \equiv \frac{|e|\mathcal{H}\hbar M}{2mc} g . \quad (74.20)$$

The factor g is called the Landé g -factor.

For singlet levels $J = L, S = 0$ we have

$$\Delta E = \frac{|e|\hbar M}{2mc}. \quad (74.21)$$

Before turning to the discussion of formulae (74.20) and (74.21) we shall establish the limits of applicability of the above derivation.

The unperturbed Pauli equation

$$\hat{H}_0 \psi = E \psi; \quad \hat{H}_0 = \frac{1}{2m} \sum_i \hat{p}_i^2 + U$$

defines the energy levels of the atom including also its multiplet structure. Thus for the theory described above to be applicable it is necessary that the matrix element of the perturbation (74.3) be smaller than the spacing between the levels corresponding to the fine structure of the atom.

It should also be pointed out that in the calculations it was assumed that Russell-Saunders coupling holds in the atom, i.e. that the quantity J is conserved in time as well as L and S .

Let us now turn to the discussion of the formulae obtained, which define the Zeeman effect.

From formula (74.20) it is seen that each component of the multiplet splits into $2J+1$ levels. Indeed, for given J the component of the total angular momentum M can take on $2J+1$ different values. This is in accordance with the result of §54: perturbation removes the degeneracy. As to the distribution of the newly arising terms, the following can be stated.

If J is an integer, then in a magnetic field a level corresponding to the value $M = 0$ arises at the place of the unperturbed level.

Of the remaining $2J$ levels J levels are distributed above and J levels below at equal distances from the level with $M = 0$.

If J is a half integer, then the levels are also distributed symmetrically with respect to the old position of the unperturbed level, and the closest levels are at the distance $|e|\hbar g/4mc$ from the initial position.

We also note that if there is jj coupling, the character of the Zeeman effect is much modified. This coupling is seldom encountered in the pure form, and we shall not carry out the corresponding calculations here.

§75. The Paschen-Back effect and the diamagnetism of atoms

In strong magnetic fields the character of the Zeeman effect changes. Namely, as the magnetic field strength increases the spacing between the

multiplet levels increases. In very strong fields the splitting of a level is so large that the spacings between the components of the multiplet arising in the field turn out to be large in comparison with the natural multiplet spacing. We recall that the latter arises from the spin-orbit interaction. In this case formula (74.20) is no longer applicable, and the character of the spectrum changes. This change in the spectrum in a strong magnetic field is called the Paschen-Back effect.

We shall carry out the calculation for the case where the splitting due to a magnetic field is large in comparison with the natural multiplet spacing. This means that the energy acquired in the magnetic field is large compared to the spin-orbit interaction. Then the term taking the spin-orbit interaction into account can be left out of the unperturbed Hamiltonian H_0 in formula (74.4). Therefore the unperturbed states of the atom can be characterized by the total angular momentum J as well as by the component L_z of the orbital angular momentum and the component S_z of the spin angular momentum.

The perturbation operator has, as before, the form

$$\hat{H}' = \frac{|e|\hbar}{2mc} (\hat{J}_z + \hat{S}_z) \mathcal{H} = \frac{|e|\hbar}{2mc} (\hat{L}_z + 2\hat{S}_z) \mathcal{H}. \quad (75.1)$$

The correction to the energy is equal to the mean value of the operator \hat{H}' over states with definite components of the orbital and spin angular momenta, i.e.

$$E' = \frac{|e|\hbar\mathcal{H}}{2mc} (\bar{L}_z + 2\bar{S}_z) = \frac{|e|\hbar\mathcal{H}}{2mc} (L_z + 2S_z). \quad (75.2)$$

Formula (75.2) defines the fine structure of the spectrum in strong magnetic fields.

Let us now consider the effect of the neglected quadratic term in the magnetic field in formula (74.2). Taking this quantity into account is particularly important for terms with $L = S = 0$. In this case no splitting of levels occurs on account of the term linear in \mathcal{H} . This can be seen from the general formula (74.20). In this case the correction due to the quadratic term cannot be disregarded. As the perturbation operator, in correspondence with formula (74.2), one has to take

$$\hat{H}'_1 = \frac{e^2}{8mc^2} \sum_i |\mathcal{H} \times \mathbf{r}_i|^2. \quad (75.3)$$

The sum over i corresponds to summing over all the electrons of the atom.

The correction to the energy levels due to the operator \hat{H}'_1 is again defined by the diagonal matrix element

$$E'_2 = \frac{e^2}{8mc^2} \sum_i \overline{|\mathfrak{A} \times \mathbf{r}_i|^2} = \frac{e^2}{8mc^2} \sum_i \overline{(\mathfrak{A} r_i \sin \theta)^2}.$$

In calculating $\overline{|\mathfrak{A} \times \mathbf{r}|^2}$ it should be recalled that the wave function of the system $L = 0, S = 0$ is spherically symmetric, hence

$$\overline{\sin^2 \theta} = 1 - \overline{\cos^2 \theta} = \frac{2}{3}.$$

Thus for the shift of the levels we obtain

$$\Delta E = \frac{e^2 \mathfrak{A}^2}{12mc^2} \sum_i \overline{r_i^2}. \quad (75.4)$$

Since the magnetic moment of the atom can be calculated by means of the formula $\mathbf{M} = -\partial \Delta E / \partial \mathfrak{A}$ (see (18.1) of Part IV), we obtain

$$\mathbf{M} = \chi \mathfrak{A}; \quad \chi = -\frac{e^2}{6mc^2} \sum_i \overline{r_i^2}. \quad (75.5)$$

Thus atoms possess diamagnetic susceptibility. Since the diamagnetic susceptibility is in the main determined by the mean square distance of the electrons from the nucleus, χ is particularly large for many-electron atoms. For such atoms good results are obtained by the Thomas-Fermi method. Hence the diamagnetic susceptibilities are often calculated by this method.

On the other hand, measurements of χ represent one of the best ways of finding the effective size of atoms. We stress that all atoms and ions have a diamagnetic susceptibility. However, in certain ions the paramagnetic susceptibility, associated with the magnetic moment, exceeds the diamagnetic susceptibility.

§76. Deuteron theory

The deuteron, consisting of a proton and a neutron, plays the same role in nuclear theory as the hydrogen atom in atomic theory.

The nuclear interaction between a proton and a neutron may depend on their separation r and the relative orientation of the spins, s_1 and s_2 , of the two particles. The explicit form of the potential energy of the nuclear inter-

action is at present unknown. Hence one has to confine oneself to writing the most general expression for the potential energy operator depending on \mathbf{r} , \mathbf{s}_1 and \mathbf{s}_2 . This interaction operator must not change under rotation of the coordinate system. Furthermore, as shown by experiment, the parity conservation law holds for nuclear forces (see §33). This means that the interaction operator must not change under reflection of coordinates (the interaction operator must commute with the parity operator). Thus we have to make up all possible scalars of the three vectors \mathbf{r} , \mathbf{s}_1 and \mathbf{s}_2 . The following scalars do not change under rotation of the coordinate system: $\mathbf{s}_1 \cdot \mathbf{s}_2$, $\mathbf{s}_1 \cdot \mathbf{r}$ and $\mathbf{s}_2 \cdot \mathbf{r}$.

The products $\mathbf{s}_1 \cdot \mathbf{r}$ and $\mathbf{s}_2 \cdot \mathbf{r}$ cannot be involved separately in the potential energy, because the spin vector is an axial vector and the product $\mathbf{s} \cdot \mathbf{r}$ is a pseudoscalar which changes sign under reflection of coordinates. The product $(\mathbf{s}_1 \cdot \mathbf{r})(\mathbf{s}_2 \cdot \mathbf{r})$ does not change sign under the reflection and, consequently, can be involved in the potential energy. Spin operators in higher powers are not involved in the operator of the interaction energy U , because the higher powers of the spin operators may be reduced by means of formula (60.17) to linear combinations of \mathbf{s} .

Thus the expression for the potential energy has the form

$$U \doteq U_1(r) + U_2(r)(\mathbf{s}_1 \cdot \mathbf{s}_2) + U_3(r)(\mathbf{s}_1 \cdot \mathbf{r})(\mathbf{s}_2 \cdot \mathbf{r}), \quad (76.1)$$

where U_1 , U_2 , U_3 are certain functions depending on the distance between the particles. Besides the operator (76.1), representing a potential energy of the ordinary type, the interaction between a proton and a neutron may also have the character of an exchange force. According to the results of §67, U_{exch} can be written by means of the exchange operator \hat{P}_{12} in the form

$$U_{\text{exch}} = \hat{P}_{12} [U_4(r) + U_5(r)(\mathbf{s}_1 \cdot \mathbf{s}_2) + U_6(r)(\mathbf{s}_1 \cdot \mathbf{r}_1)(\mathbf{s}_2 \cdot \mathbf{r}_2)]. \quad (76.2)$$

Here U_4 , U_5 and U_6 are functions of the distance between the particles independent of their spins. For generality it is assumed that the form of these functions is different from the form of the functions U_1 , U_2 , U_3 involved in the potential energy of the ordinary interaction. The total interaction energy is equal to the sum of expressions (76.1) and (76.2). Available data on the stable states of the deuteron, the study of neutron-proton scattering etc. do not, as yet, allow one to determine the form of these functions. Moreover, there are no grounds for considering any of these functions to be small in comparison with the others. Thus even the simplest nuclear system turns out to be immeasurably more complex than atomic systems.

Experimental data already makes it possible to carry out a classification of the states of the deuteron. As is easily seen, the Hamiltonian of a system of two nucleons (a proton and a neutron) with the interaction energy written

above leads to two conservation laws: the total angular momentum conservation law and the parity conservation law.

The states of the deuteron are denoted by the same symbols as the states of atoms. States with the orbital angular momentum $L = 0, 1, 2, \dots$ are denoted respectively by S, P, D and so on. The multiplicity of the $(2S+1)$ th term is denoted by a superscript on the left (S is the total spin of the deuteron). The subscript on the right denotes the total angular momentum J of the deuteron. For example, in the state 3P_0 the total spin is equal to one, $L = 1$, and the total angular momentum is equal to zero.

Let us discuss the possible states of the system taking into account the fact that the spins of the neutron and proton are equal to $\frac{1}{2}$. Formal application of the rule of addition of angular momenta leads to the following possible states of the system:

$$\begin{aligned} & {}^1S_0, \quad {}^1P_1, \quad {}^1D_2 \quad (\text{singlets}), \\ & {}^3S_1, \quad {}^3P_0, \quad {}^3P_1, \quad {}^3P_2, \quad {}^3D_1, \quad {}^3D_2, \quad {}^3D_3 \quad (\text{triplets}). \end{aligned}$$

The S- and D-states are even, whereas the P-state is odd. We have not written down states with $L > 2$. The states realized in nature can be deduced only from experimental data. Experiment shows that the ground state of the deuteron is an even state with $J = 1^*$.

Further, making use of the rules of addition of angular momenta, we shall establish the possible states of a system with total angular momentum $J = 1$.

The total spin of a system consisting of a neutron and a proton can be equal either to zero or to one. If the spin is equal to zero, then only one state $L = 1$ leading to the total angular momentum $J = 1$ is possible. For a spin equal to one the orbital angular momentum can take on three values: $L = 0, 1, 2$. Consequently, four states are in all possible: ${}^1P_1, {}^3S_1, {}^3P_1, {}^3D_1$; the states 1P_1 and 3P_1 cannot be realized, because they are odd.

Further, it is easily seen that superpositions of states such as ${}^3S_1 + {}^3P_1$ or ${}^3S_1 + {}^1P_1$ are impossible, since S- and P-states have different parities and the wave function corresponding to their superposition is not an eigenfunction of the parity operator.

Thus the deuteron can be either in the state 3S_1 or in the state 3D_1 or in a state which is a superposition of these two states. The S-state is spherically

* See L.D.Landau and Ya.Smorodinskii, *Lectures on nuclear theory* (Consultants Bureau, New York, 1958); A.I.Akhiezer and I.Ya.Pomeranchuk, *Nekotorye voprosy teorii yadra* (Some problems of nuclear theory) (Gostekhizdat, Moscow, 1950) §2 and §5.

symmetric. If the deuteron were in this state, then its quadrupole moment would be equal to zero. Experiment shows, however, that the quadrupole moment of the deuteron is different from zero, although it is small. This means that the normal state of the deuteron represents the superposition of the spherically symmetric 3S_1 -state and the asymmetric 3D_1 -state.

Knowing the experimental value of the quadrupole moment of the deuteron, one can estimate the contribution given by the 3D_1 -state to the wave function of the deuteron. This contribution turns out to be small. Thus it can be assumed that the deuteron is a spherically symmetric system with a small admixture of asymmetry brought about by the D-state.

We shall consider further a rough model of the deuteron in which we assume that the potential energy of the interaction between the neutron and the proton depends only on the distance between them. In other words, we shall retain only the first term $U_1(r) \equiv U(r)$ in formula (76.1). We shall disregard the asymmetry of the deuteron, assuming it to be in the ground state. The equation for the relative motion of the neutron and proton can be written, in correspondence with formula (14.11), in the form

$$\left[-\frac{\hbar^2}{2\mu} \nabla^2 + U(r) \right] \psi_0 = \epsilon \psi_0. \quad (76.3)$$

In this case the reduced mass of the system is equal to

$$\frac{1}{\mu} = \frac{1}{m_p} + \frac{1}{m_n},$$

where m_p is the mass of the proton, and m_n is the mass of the neutron.

Since $m_p \approx m_n$, we obtain

$$\frac{1}{\mu} = \frac{2}{m_p}.$$

As to the potential energy $U(r)$ we confine ourselves only to the general assumption that it tends to 0 rapidly as $r \rightarrow r_0$, where r_0 is the range of nuclear forces. We cannot give the concrete form of $U(r)$ for $r < r_0$ since we do not know the law of interaction of nuclear forces.

If the function ψ_0 is sought in the form

$$\psi_0(r) = \frac{\chi(r)}{r}, \quad (76.4)$$

then making use of formula (35.16) with $l = 0$ we obtain the equation for the function $\chi(r)$

$$\left[-\frac{\hbar^2}{m_p} \frac{d^2}{dr^2} + U(r) \right] \chi(r) = \epsilon \chi(r). \quad (76.5)$$

For $r > r_0$ eq. (76.5) is written in the form

$$-\frac{\hbar^2}{m_p} \frac{d^2 \chi}{dr^2} = \epsilon \chi. \quad (76.6)$$

We seek the solution decreasing at infinity in the form

$$\chi = C e^{-\alpha r}. \quad (76.7)$$

Substituting (76.7) into (76.6) we obtain the relation for α

$$-\frac{\hbar^2}{m_p} \alpha^2 = \epsilon = -|\epsilon|, \quad \alpha = \left[\frac{m_p |\epsilon|}{\hbar^2} \right]^{\frac{1}{2}}. \quad (76.8)$$

Then for the wave function we have

$$\psi_0 = C \frac{e^{-\alpha r}}{r}. \quad (76.9)$$

As a characteristic of the size of the deuteron one can choose the quantity $r_1 = \alpha^{-1}$, i.e. the distance at which the wave function χ decreases by a factor e . The distance r_1 is easily determined from relation (76.8), since the binding energy of the deuteron is well known from experimental data to be $|\epsilon| = 2.19$ MeV. Substituting the value of \hbar and m_p into formula (76.8), we obtain $r_1 = 4.3 \times 10^{-13}$ cm. Consequently, the wave function ψ_0 of the deuteron differs from zero in a range considerably larger than the range of nuclear forces ($r_0 \approx 2 \times 10^{-13}$ cm). Thus we see that the neutron and proton can be observed with a high probability at distances from each other which substantially exceed the size of the sphere of action of nuclear forces.

The dependence of the wave function ψ_0 on the distance cannot be determined in the region $r < r_0$, since the potential energy in this region is unknown.

However, from the general theory of motion in a spherically symmetric field it follows that for $r \rightarrow 0$ the function χ is proportional to r^{l+1} (see §35) and, consequently, in the S-state is proportional to r . Thus at small distances the function χ tends to zero.

The constant C contained in the ψ -function can be found from the normalization condition. For the wave function ψ_0 for $r < r_0$, we take it in the form (76.9), which we assume to be valid over all space. This does not introduce a substantial error, since a large part of the normalization integral

refers to the region $r > r_0$. Substituting (76.9) into the normalization condition, we find

$$C = (\alpha/2\pi)^{\frac{1}{2}}. \quad (76.10)$$

Let us now establish the general relation between the width of the well r_0 and its depth. For this we integrate eq. (76.5) in the range from zero to $r = r_0$. As a result of the integration we obtain

$$\chi'_{r=r_0} - \chi'_{r=0} = \frac{m_p}{\hbar^2} \int_0^{r_0} U(r)\chi(r) dr + \frac{m_p|\epsilon|}{\hbar^2} \int_0^{r_0} \chi dr. \quad (76.11)$$

As can be seen from fig. V.22, the value of the derivative $|\chi'|$ taken at the point $r = r_0$ is considerably smaller than that of the derivative $|\chi'_{r=0}|$. Furthermore, one can disregard the binding energy in comparison with the potential energy of interaction, i.e. one can assume that $|\epsilon| < |U(r)|$ for $r < r_0$. At small distances $\chi = Nr$, where N is a constant. Then (76.11) transforms into the form

$$-N = \frac{m_p}{\hbar^2} N \int_0^{r_0} U(r)r dr \quad (76.12)$$

or

$$\int_0^{r_0} U(r)r dr = -\frac{\hbar^2}{m_p}. \quad (76.13)$$

Replacing the integral in (76.13) by $U_0 r_0^2$, where U_0 is a mean energy of interaction, i.e. the mean depth of the well, we obtain in order of magnitude

$$U_0 \approx -\frac{\hbar^2}{m_p r_0^2} \sim -40 \text{ MeV}.$$

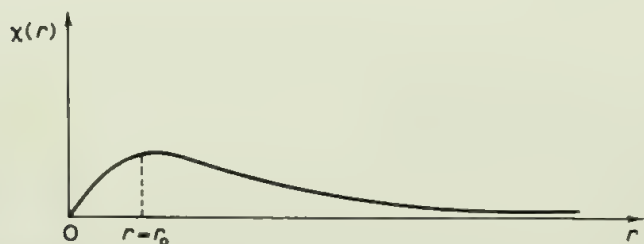


Fig. V.22

§77. Nuclear-shell theory

In contrast to atoms, in which the interaction between electrons is of a secondary character and takes place against the background of the principal interaction (the attraction towards the nucleus) there is no single centre of interaction in atomic nuclei.

On the contrary, all the nuclear particles (nucleons) interact intensively with each other through strong nuclear forces. Hence for a long time it seemed that it made no sense to distinguish between the states of the individual particles in the nucleus, and that one could speak only of the state of the system as a whole.

It turned out, however, that a number of the observed properties of atomic nuclei pointed to the conservation of the individuality of nucleons in nuclei. Apparently, the conservation of individuality of particles in nuclei is associated with the fact that the nuclear forces decrease very rapidly with increasing distance, and that the kinetic energy of the nucleons in nuclei is very large.

Proceeding from the assumption that each of the nucleons in a nucleus moves in the self-consistent field produced by all the other nucleons, it was possible to account for a number of important properties of nuclei.

The self-consistent field of most nuclei is spherically symmetric. Of course, the precise law of the potential distribution inside the nucleus is unknown. It turns out, however, that the character of the distribution of levels depends relatively little on the model of the potential field adopted, provided it gives correctly the basic feature of the field of the nucleus as a whole: the sharp increase of the potential at its surface $r = R$. The simplest nuclear model is a spherical potential well of infinite depth. In this model the self-consistent potential field U in which an individual nucleon moves has the form

$$U = \begin{cases} 0 & \text{for } r < R \\ \infty & \text{for } r \geq R \end{cases}.$$

Already this very simplified model allows one to get a general idea of the properties and distribution of levels. The wave function satisfies the equation

$$\frac{d^2\psi}{dr^2} + \frac{2}{r} \frac{d\psi}{dr} + \frac{l(l+1)}{r^2} \psi - k^2 \psi = 0, \quad (77.1)$$

where $k^2 = 2mE/\hbar^2$. The solution of this equation is expression (36.9). The boundary condition at the surface of the nucleus

$$\psi = 0 \quad \text{for } r = R \quad (77.2)$$

leads to the condition

$$J_{l+\frac{1}{2}}(kR) = 0. \quad (77.3)$$

For a given value of the orbital angular momentum the energy levels, which are the roots of the transcendental equation (77.3), are classified by means of the principal quantum number n . To the smallest root of this transcendental equation there corresponds a wave function having no nodes for $r < R$. This level is classified as a level with $n = 1$. The next root of (77.3) corresponds to a wave function having one node for $r < R$. The corresponding state is denoted by $n = 2$ and so on. For a given n , l can take on any value. The energy of the nucleon increases with increasing orbital angular momentum.

The ordering of levels is given by the sequence

$$1s, 1p, 1d, 2s, 1f, 2p, 1g, 2d \dots$$

It turned out that in atomic nuclei an important role is played by the spin-orbit interaction, which has so far not been taken into account at all. The special role of the spin-orbit interaction is associated with the fact that, owing to the rapid decrease of nuclear forces with increasing distance between the particles, the energy of pair interaction is on the average small. The considerable magnitude of the spin-orbit interaction leads to the establishment of jj coupling in nuclei. The spin and orbital angular momenta of each nucleon are added up into the total angular momentum j . The energy of a nucleon turns out to depend on its spin. In nuclei there exists the inverted structure of levels, in which levels with large j lie below those with smaller j . The states of nucleons are denoted by the symbol nl_j ; for example $1s_{1/2}$ or $2p_{3/2}$. Since nucleons obey the Pauli principle, $2j+1$ neutrons and $2j+1$ protons can be in each state with given values of n , l and j . Because of this there is a situation in nuclei very similar to that in atoms: the states of the nucleons can be divided into groups or shells. When each shell is filled a closed configuration arises, possessing the greatest stability — in the given case the largest energy of binding of the nucleon in the nucleus. The energy distribution of the states and the number of nucleons in these states is given by the table

$1s_{1/2}$						2
$1p_{3/2},$	$1p_{1/2}$					6
$1d_{3/2},$	$1d_{1/2},$	$2s_{1/2},$	$1f_{1/2}$			20
$2p_{3/2},$	$1f_{5/2},$	$2p_{1/2},$	$1g_{3/2}$			22
$2d_{5/2},$	$1g_{7/2},$	$1h_{11/2},$	$2d_{3/2},$	$3s_{1/2}$		32
$2f_{7/2},$	$1h_{9/2},$	$2i_{13/2},$	$2f_{3/2},$	$3p_{3/2},$	$3p_{1/2}$	44

From this table it is seen that in the closed shells there are consecutively 2, 8, 28, 50, 82 and 126 nucleons.

Correspondingly, nuclei with the total number of nucleons given by these numbers, which are called magic numbers, possess particular stability.

Nuclei in which both the number of protons and the number of neutrons are magic numbers are particularly stable. Such nuclei are, for example, ${}^4_2\text{He}$ and ${}^{16}_8\text{O}$. They are often said to be doubly magic. This very simple scheme makes it possible to account not only for a particular stability and natural abundance of certain isotopes but also a number of other properties of atomic nuclei, for example their magnetic moments. However, in a number of cases it turns out to be inadequate. Thus, for example, nuclei with open shells display deviations from spherical form. This is exhibited in the presence of rotation of the nucleus as a whole. The experimental proof of rotation of the nucleus is the presence in nuclear spectra of a structure similar to that of the spectra of diatomic molecules.

We cannot dwell on details here, but the reader may find these in the specialized literature*.

* See, for example, P.E.Nemirovskii, *Contemporary models of the atomic nucleus* (Pergamon Press, Oxford, 1963).

The Theory of Diatomic Molecules

§78. The adiabatic approximation and the classification of electron terms

We now turn to the study of the properties of a more complex system, i.e. a molecule. We shall confine ourselves to the consideration of the simplest diatomic molecule. The basic properties of molecular systems will be illustrated in this simple example.

We have seen in the preceding chapter that the calculation for atoms is carried out by approximate methods. It is natural that approximate methods of calculation should also be widely used in the theory of molecules. The Hamiltonian of the diatomic molecule is of the form

$$\left[-\frac{\hbar^2}{2M_1} \nabla_1^2 - \frac{\hbar^2}{2M_2} \nabla_2^2 - \sum_{k=1}^n \frac{\hbar^2}{2m} \nabla_k^2 + U(\mathbf{r}_k, \mathbf{R}_i) \right] \psi_{\text{mol}}(\mathbf{r}_k, \mathbf{R}_i) =$$

$$= E \psi_{\text{mol}}(\mathbf{r}_k, \mathbf{R}_i) . \quad (78.1)$$

Here M_1 and M_2 are the masses of the nuclei, m is the mass of an electron, \mathbf{r}_k are the coordinates of the electrons, and \mathbf{R}_i are the coordinates of the nuclei. The potential energy $U(\mathbf{r}_k, \mathbf{R}_i)$ involves the interaction of the electrons with the nuclei, the interaction of the electrons with each other and the interaction of the nuclei with each other. The summation over k is carried out over all the

electrons of the molecule. The velocities of the nuclei, which have masses larger than the electron mass by a factor of several thousand, are substantially lower than the velocities of the electrons. Correspondingly, in the theory of molecules use is made of the adiabatic approximation (cf. §57). The wave function of the system is written in the form

$$\psi_{\text{mol}} = \alpha_n(\mathbf{R}_i) \psi_n(\mathbf{r}_k, \mathbf{R}_i).$$

Using formulae (57.7) and (57.8), we can write the equations for the functions α and ψ :

$$\left[-\frac{\hbar^2}{2m} \sum \nabla_k^2 + U(\mathbf{r}_k, \mathbf{R}_i) \right] \psi_n = E_n(\mathbf{R}_i) \psi_n, \quad \hat{H}_{\text{el}} \psi_n = E_n \psi_n, \quad (78.2)$$

$$\left[-\frac{\hbar^2}{2M_1} \nabla_1^2 - \frac{\hbar^2}{2M_2} \nabla_2^2 + E_n(\mathbf{R}_i) \right] \alpha_n(\mathbf{R}_i) = E \alpha_n(\mathbf{R}_i). \quad (78.3)$$

Eq. (78.2) describes the motion of electrons for nuclei at rest. The quantity $E_n(\mathbf{R}_i)$ defines the energy levels of the system for motionless nuclei which are at a fixed distance from each other. The energy E_n for a fixed distance between the nuclei is called the electron term.

The Schrödinger equation (78.3) describes the motion of the nuclei. The quantity $E_n(\mathbf{R}_i)$ in it is the potential energy of the nuclei. From eq. (78.3) we see that the total energy of the nuclei E depends on the state of the electron part of the system, i.e. on the electron term.

The number of electrons in a molecule is always greater than one. Hence even the solution of the approximate equation (78.2) is associated with great mathematical difficulties which are insuperable in the case of many-electron molecules. (An exception to this is the ion H_2^+ , for which a precise solution of the Schrödinger equation for the electron part of the wave function has been obtained.) We are forced, without trying to solve eq. (78.2), to find the most general properties of a system of electrons moving in the field of two nuclei.

For this purpose we find, as usual, quantities which commute with the Hamiltonian \hat{H}_{el} , in other words we find quantities which simultaneously have definite values in stationary states of the system. In contrast to the atomic field, which possesses spherical symmetry, the field of a diatomic molecule has cylindrical symmetry. The symmetry axis is the straight line joining the two nuclei, which we shall choose to be the z -axis in what follows. The potential energy of the interaction of the electrons with the nuclei, as well as that of the interaction of the electrons with each other, does not change under a rotation through the angle φ with respect to the z -axis. Hence the Hamiltonian of the system of electrons

$$\hat{H}_{\text{el}} = -\frac{\hbar^2}{2m} \sum_k \nabla_k^2 + U$$

does not depend on the angle φ .

Thus we arrive at the conclusion that the component of the total angular momentum of the electrons along the axis of the molecule is conserved. Disregarding the weak spin-orbit interaction, it can be assumed that the component of the orbital angular momentum of the electrons in the z -direction is also conserved. The states of the electrons are classified according to the eigenvalues of the operator L_z . The eigenvalues of the z -component of the orbital angular momentum of the electrons are denoted by the letter Λ . States with $\Lambda = 0, 1, 2$ are called Σ -, Π - and Δ -states (in analogy with the S-, P- and D-states of atoms). The total spin of the electrons S is also conserved for a system of electrons in a molecule.

All the reasoning about the total spin which was presented in the theory of the atom applies also to the molecule.

As in the case of an atom, the multiplicity $2S+1$ of an electron term of a molecule is indicated in the form of a superscript on the left of the quantum number Λ , i.e. in the form $^{2S+1}\Lambda$.

Further, we shall show that the Hamiltonian \hat{H}_{el} does not change under the reflection of the coordinates of the electrons in any plane passing through the nuclei of the molecule (i.e. through the z -axis). In other words, we shall show that the Hamiltonian commutes with the reflection operator. This can easily be seen if, for example, the plane in which the reflection takes place is chosen in such a way that it passes through the z -axis and y -axis. In this case the reflection corresponds to the replacement of all coordinates $x_i \rightarrow -x_i$. But, since the interaction depends only on the distance between the particles $(x_1 - x_2)^2$, it becomes evident that the Hamiltonian commutes with the reflection operator. Hence it follows that the operator \hat{H}_{el} and the reflection operator in a plane passing through the axis of the molecule have common eigenfunctions. Therefore stationary states can be characterized, in addition to the eigenvalues Λ , by the eigenvalues P_i of the reflection operator. The latter, as is easily seen, takes on two values $P_i = \pm 1$. However, things are complicated by the fact that the angular momentum component operator \hat{L}_z does not commute with the reflection operator. Indeed, the operator of the z -component of the angular momentum has the form

$$\hat{L}_z = -i\hbar y \frac{\partial}{\partial x} + i\hbar x \frac{\partial}{\partial y}.$$

The x -coordinate will change sign under reflection in the zy -plane, whereas

the y -coordinate will not. Hence it follows directly that the operators \hat{P}_i and \hat{L}_z do not commute. Therefore molecular terms cannot be simultaneously characterized by means of the quantities Λ and P_i , except for terms for which $L_z = 0$. In this last case, states with parity $P_i = 1$ and $P_i = -1$, which are denoted by Σ^+ and Σ^- , are possible.

The wave functions corresponding to these states change and do not change sign respectively under the action of the operator \hat{P}_i , corresponding to reflection in a plane passing through the nuclei of the molecule.

Let us now consider the particular case of a molecule with identical nuclei. If the origin is chosen to be at the point which lies on the z -axis halfway between the nuclei, then it is easily seen that the operator of inversion of the electron coordinates (corresponding to the replacement of all coordinates of the electrons by the inverse coordinates $\mathbf{r}_i \rightarrow -\mathbf{r}_i$) commutes with the Hamiltonian \hat{H}_{el} . Since at the same time the operator \hat{P}_i of the reflection of electron coordinates in a plane passing through the nuclei of the molecule commutes with \hat{H}_{el} , the state with $L = 0$ can be characterized by three eigenvalues $\Lambda = 0$, $P_i = \pm 1$ and the eigenvalues of the inversion operator (see §33). The latter has two values ± 1 denoted by the letters g (even state) and u (odd state). These indices are written as subscripts on the right. For example, $^1\Sigma_g^+$ corresponds to a term whose wave function is even and does not change sign under the action of the operator of reflection in a plane passing through the z -axis; the z -component of the angular momentum is equal to zero; the term is singlet. Further, we know that the inversion operator commutes with the operator L_z . Hence the states Π and Δ also can be both even and odd. In other words, the states Π_u , Π_g , Δ_u , Δ_g and so on are possible.

Let us dwell on the problem of degeneracy of the electron terms. If Λ is defined, then this means that the absolute value of the z -component of the angular momentum is defined. Since the energy of the system cannot depend on the orientation of the angular momentum component with respect to the z -axis, i.e. is the same for $L_z = +\Lambda$ and $L_z = -\Lambda$, we arrive at the conclusion that each term with $L_z \neq 0$ is two-fold degenerate. Finally, we point out that the energy of the electron terms of the molecule is of the same order as the energy of atomic terms.

§79. The hydrogen molecule. Ideas of the theory of chemical binding

The only molecule for which one can obtain a reasonably accurate solution of the equation for the electron term is the hydrogen molecule. This calculation is of great theoretical importance.

If the energy is measured with respect to the energy of the separated motionless atoms, then the energy values of the electron terms of the stable molecule have negative values. The energy (negative) of the molecule is a measure of the chemical binding of the constituent atoms. Thus the calculation of the electron terms of the molecule represents at the same time a quantitative theory of chemical binding between atoms.

The establishment of the nature of chemical binding is one of the fundamental results of quantum mechanics.

Before the appearance of quantum mechanics there were no substantiated concepts of the nature of chemical binding, in particular of the nature of homopolar molecules. We recall that homopolar molecules are molecules made up of neutral atoms. For example, molecules containing identical atoms are of this type. We shall try to account for certain characteristic features of the theory of chemical binding in the example of the hydrogen molecule.

The Schrödinger equation for the electron terms of the hydrogen molecule is of the form

$$\left(-\frac{\hbar^2}{2m} \nabla_1^2 - \frac{\hbar^2}{2m} \nabla_2^2 + \frac{e^2}{R} - \frac{e^2}{r_{a1}} - \frac{e^2}{r_{a2}} - \frac{e^2}{r_{b1}} - \frac{e^2}{r_{b2}} + \frac{e^2}{r_{12}} \right) \psi = E\psi. \quad (79.1)$$

Here R is the distance between the nuclei of the hydrogen atoms; the quantities r_{a1} , r_{a2} , r_{b1} , r_{b2} are respectively the distances between nucleus a and the first electron, nucleus a and the second electron, nucleus b and the first electron, and nucleus b and the second electron; and r_{12} is the distance between the electrons.

If the atoms forming the molecule are placed an infinitely large distance apart, then it can be said that one of the electrons, for example N_1 , will be bound to nucleus a , and the other (electron N_2) to nucleus b . By virtue of the identity of electrons, such a statement makes no sense when the atoms are brought together.

An exact solution of the Schrödinger equation (79.1) involves great mathematical difficulties. Hence a number of approximate calculations have been carried out. We shall make use of perturbation theory, which allows the basic properties of the system to be elucidated relatively simply. The question as to the degree of accuracy of the calculation will be discussed later. We choose as the wave function of zero order approximation the wave function of the system with infinitely distant nuclei. For an infinite distance between the nuclei ($R \rightarrow \infty$) the wave function of the two electrons and two nuclei has the form

$$\varphi_1^0 = \psi_a(\mathbf{r}_{a1})\psi_b(\mathbf{r}_{b2}), \quad (79.2)$$

where $\psi_a(\mathbf{r}_{a_1})$ and $\psi_b(\mathbf{r}_{b_2})$ are the wave functions of the hydrogen atom in which respectively the first electron is near nucleus a and the second electron is near nucleus b .

Evidently, these functions satisfy the equations

$$\begin{aligned} \left(-\frac{\hbar^2}{2m} \nabla_1^2 - \frac{e^2}{r_{a_1}}\right) \psi_a(\mathbf{r}_{a_1}) &= E_0 \psi_a(\mathbf{r}_{a_1}), \\ \left(-\frac{\hbar^2}{2m} \nabla_2^2 - \frac{e^2}{r_{b_2}}\right) \psi_b(\mathbf{r}_{b_2}) &= E_0 \psi_b(\mathbf{r}_{b_2}). \end{aligned} \quad (79.3)$$

We are interested in the normal state of the hydrogen molecule. Therefore E_0 must be understood to be the lowest energy level of the hydrogen atom. Indeed, the same energy is possessed by the state

$$\varphi_2^0 = \psi_a(\mathbf{r}_{a_2}) \psi_b(\mathbf{r}_{b_1}), \quad (79.4)$$

which differs from the first state by electron exchange. We stress that φ_1^0 and φ_2^0 are eigenfunctions of different operators and are not orthogonal to each other.

We write the wave functions of the zero order approximation in the form of symmetrized combinations of the functions φ_1^0 and φ_2^0 , i.e. as

$$\begin{aligned} \psi_s^0 &= A_1 [\psi_a(\mathbf{r}_{a_1}) \psi_b(\mathbf{r}_{b_2}) + \psi_a(\mathbf{r}_{a_2}) \psi_b(\mathbf{r}_{b_1})], \\ \psi_a^0 &= A_2 [\psi_a(\mathbf{r}_{a_1}) \psi_b(\mathbf{r}_{b_2}) - \psi_a(\mathbf{r}_{a_2}) \psi_b(\mathbf{r}_{b_1})]. \end{aligned} \quad (79.5)$$

The constants A_1 and A_2 are defined by the normalization condition

$$\int |\psi_s^0|^2 dV_1 dV_2 = \int |\psi_a^0|^2 dV_1 dV_2 = 1.$$

They are equal to

$$A_1 = [2(1+s^2)]^{-\frac{1}{2}}, \quad A_2 = [2(1-s^2)]^{-\frac{1}{2}},$$

where the quantity s represents the degree of non-orthogonality of the functions φ_1^0 and φ_2^0 and is equal to

$$s^2 = \int \varphi_1^0 \varphi_2^0 dV_1 dV_2. \quad (79.6)$$

For such a choice of unperturbed functions, which represent the symmetrized wave functions of individual atoms, the perturbation operator is expressed by

$$\hat{H}' = \frac{e^2}{R} - \frac{e^2}{r_{a_2}} - \frac{e^2}{r_{b_1}} + \frac{e^2}{r_{12}}.$$

Direct application of perturbation theory is inadmissible: the zero order wave functions are not orthogonal to each other. Therefore it is necessary to modify the perturbation theory somewhat. We write the perturbed wave functions and the energy of the perturbed system in the form

$$\begin{aligned}\psi_s &= \psi_s^0 + \psi'_s, \\ \psi_a &= \psi_a^0 + \psi'_a, \\ E &= E_0 + \epsilon.\end{aligned}\tag{79.7}$$

Then eq. (79.1) for the function ψ'_s is written in the form

$$\begin{aligned}\frac{\hbar^2}{2m}(\nabla_1^2 + \nabla_2^2)A_1[\psi_a(r_{a_1})\psi_b(r_{b_2}) + \psi_a(r_{a_2})\psi_b(r_{b_1})] + \frac{\hbar^2}{2m}(\nabla_1^2 + \nabla_2^2)\psi'_s + \\ + \left[E_0 + \epsilon - \frac{e^2}{R} - \frac{e^2}{r_{12}} + \frac{e^2}{r_{a_1}} + \frac{e^2}{r_{a_2}} + \frac{e^2}{r_{b_1}} + \frac{e^2}{r_{b_2}}\right] \times \\ \times A_1[\psi_a(r_{a_1})\psi_b(r_{b_2}) + \psi_a(r_{a_2})\psi_b(r_{b_1})] + \\ + \left[E_0 + \epsilon - \frac{e^2}{R} - \frac{e^2}{r_{12}} + \frac{e^2}{r_{a_1}} + \frac{e^2}{r_{a_2}} + \frac{e^2}{r_{b_1}} + \frac{e^2}{r_{b_2}}\right]\psi'_s = 0.\end{aligned}$$

Using eq. (79.3) and dropping small terms containing the product of the perturbation operator with the perturbed wave function, we obtain

$$\begin{aligned}\left\{\frac{\hbar^2}{2m}(\nabla_1^2 + \nabla_2^2) + \frac{e^2}{r_{a_1}} + \frac{e^2}{r_{b_2}} + E_0\right\}\psi'_s = \\ = -\left\{A_1\left[\epsilon - \frac{e^2}{R} - \frac{e^2}{r_{12}}\right]\psi_s^0 + A_1\left(\frac{e^2}{r_{b_1}} + \frac{e^2}{r_{a_2}}\right)\varphi_1^0 + \right. \\ \left. + A_1\left(\frac{e^2}{r_{a_1}} + \frac{e^2}{r_{b_2}}\right)\varphi_2^0\right\}.\end{aligned}\tag{79.8}$$

An analogous expression is obtained upon substituting ψ_a .

For further calculations we shall make use of the following general theorem. In order that the solution of the equation with the right-hand side

$$(\hat{H}_0 - E_k^0)\psi = \varphi$$

may exist it is necessary that the right-hand side φ be orthogonal to the function ψ_k^0 satisfying a homogeneous equation.

(The proof is particularly easy for the case of a non-degenerate spectrum. We write the equation in the form

$$(\hat{H}_0 - E_n^0)\psi = \varphi, \quad (79.9)$$

where \hat{H}_0 is a linear operator having the non-degenerate spectrum of eigenvalues

$$\hat{H}_0\psi_k^0 = E_k^0\psi_k^0.$$

We expand the function ψ in terms of the functions ψ_k^0 :

$$\psi = \sum a_k \psi_k^0.$$

Substituting into (79.9), we have

$$\sum_{k \neq n} a_k (E_k^0 - E_n^0) \psi_k^0 = \varphi.$$

Calculating the integral $\int \psi_n^0 \varphi dx$, we find that it is equal to zero. This proves our statement.)

Applying this theorem to (79.8), we require that the right-hand side of (79.8) be orthogonal to the solution of the homogeneous equation, i.e. to the unperturbed eigenfunction φ_1^0 . Multiplying the right-hand side of (79.8) by φ_1^0 and integrating, we find

$$\begin{aligned} \int \left(\epsilon - \frac{e^2}{r_{12}} - \frac{e^2}{R} \right) [\varphi_1^0 \pm \varphi_2^0] \varphi_1^0 dV + \\ + \int \left(\frac{e^2}{r_{b_1}} + \frac{e^2}{r_{a_2}} \right) (\varphi_1^0)^2 dV \pm \int \left(\frac{e^2}{r_{a_1}} + \frac{e^2}{r_{b_2}} \right) \varphi_1^0 \varphi_2^0 dV = 0, \end{aligned}$$

where the plus sign refers to the symmetric function. Analogous calculations lead to a formula with minus sign in the case of the antisymmetric function.

Taking into account the normalization condition, we have

$$\begin{aligned} \epsilon(1 \pm s^2) = \int \left(\frac{e^2}{R} + \frac{e^2}{r_{12}} - \frac{e^2}{r_{b_1}} - \frac{e^2}{r_{a_2}} \right) (\varphi_1^0)^2 dV \pm \\ \pm \int \left(\frac{e^2}{R} + \frac{e^2}{r_{12}} - \frac{e^2}{r_{a_1}} - \frac{e^2}{r_{b_2}} \right) \varphi_1^0 \varphi_2^0 dV \end{aligned}$$

Solving this equation for ϵ , we find the correction to the energy

$$\epsilon = \frac{J \pm K}{1 \pm s^2}, \quad (79.10)$$

where

$$\begin{aligned}
 J(R) &= e^2 \int \psi_a^2(\mathbf{r}_{a_1}) \psi_b^2(\mathbf{r}_{b_2}) \left(\frac{1}{R} - \frac{1}{r_{a_2}} - \frac{1}{r_{b_1}} + \frac{1}{r_{12}} \right) dV_1 dV_2, \\
 K(R) &= e^2 \int \psi_a(\mathbf{r}_{a_1}) \psi_b(\mathbf{r}_{b_1}) \psi_a(\mathbf{r}_{a_2}) \psi_b(\mathbf{r}_{b_2}) \times \\
 &\quad \times \left(\frac{1}{R} - \frac{1}{r_{a_1}} - \frac{1}{r_{b_2}} + \frac{1}{r_{12}} \right) dV_1 dV_2.
 \end{aligned}
 \tag{79.11}$$

In formula (79.10) the plus sign corresponds to the symmetric state characterized by the wave function ψ_s^0 , and the minus sign corresponds to the antisymmetric state (the wave function ψ_a^0). If ψ_s^0 and ψ_a^0 were mutually orthogonal, then we would find that $s = 0$. In this case (79.10) would be the same as the ordinary formula of perturbation theory for the correction to the energy.

The expressions found for J and K are analogous to the integrals obtained in §67. The integral J defines the Coulomb interaction of the nuclei and electrons with each other. The quantity K represents the exchange energy. The total wave function of the hydrogen molecule represents the product of the spatial functions and the spin function. Since the total wave function of the system must be antisymmetric, the spatial function symmetric in coordinates must be multiplied by the antisymmetric spin function and vice versa. Hence the state (ψ_s^0) is a state with zero spin for the electrons (singlet state). Analogous reasoning shows that the wave function ψ_a^0 describes a state of the molecule with a spin of one. To find out which of these two states is the bound state (molecule), it is necessary to find the dependence of the quantity ϵ on the radius R . This can be done if the wave functions of the normal state of the hydrogen atom are substituted for ψ_a and ψ_b in integral (79.11). The results of the calculation are conveniently presented in a graph (fig. V.23). Here E_1 and E_2 are the energies of the molecule corresponding respectively to the singlet and triplet states. We see that two hydrogen atoms having the total electron spin equal to one cannot form a bound state, since E_2 has no minimum. The bound state can only be the singlet state. Knowing the form of $E_1(R)$, it is possible to determine the binding energy as well as the effective size of the molecule. The minimum of the potential energy lies at $R_0 = 0.79 \text{ \AA}$. The binding energy is not in very good agreement with experimental values. This is due to the fact that the operator chosen as the perturbation does not contain a small parameter and is not small in comparison with the unperturbed operator. Therefore quantitative application of the perturbation theory is inapplicable. Somewhat better results are obtained by other approximate methods. However, the general qualitative conclusions on the

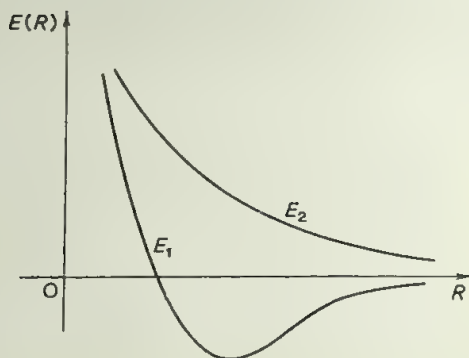


Fig. V.23

nature of chemical binding with the formation of a stable molecule from two atoms is correct. The stability of the molecule is wholly determined by the magnitude and sign of the exchange integral. Namely, for a stable chemical compound to be formed it is necessary (although not sufficient) that the spins of the atoms be antiparallel. This is often expressed by the not strictly correct statement that 'the forces of chemical binding are exchange forces'. In §67 we have discussed in detail the quantum-mechanical theory of exchange forces and the meaning of such formulations. Anyhow, it is beyond doubt that the forces responsible for the formation of homopolar chemical compounds are of a specific quantum-mechanical character. It is also often said that 'antiparallel spins are coupled'. The preceding calculation clearly shows the large degree of relativity of such a terminology. In the example of the hydrogen molecule one can show not only the quantum-mechanical nature of the forces of chemical binding but also the difficulties arising in calculating the formation of molecular systems.

The above calculation allows one to draw the general conclusion that to each valence in a chemical compound there corresponds a pair of electrons with antiparallel spins, bound to each other by the exchange interaction.

Hence it follows that unpaired outer electrons, which are in open shells, are responsible for the chemical properties of atoms. The valence of the atom is determined by the number of such unpaired electrons. When a homopolar chemical compound is formed these electrons get 'collectivized', i.e. can no longer be considered to belong to a given atom. At the same time the configuration of the unfilled state is modified in such a way that it approaches a filled structure. In other words, the electrons get paired in such a way that

the spins of all the electrons in the molecule tend to compensate for each other. Stable homopolar molecules tend to have all their electrons paired. In the most stable molecules all electrons are paired, the total spin of the molecule is $S = 0$ and the multiplicity is equal to one.

These qualitative results are in good agreement with experimental data for most molecules. We should also mention some facts which are important for understanding the formation of molecules. The first of these is the property (mentioned in §73) of atoms to enter into chemical binding in an excited state rather than in the normal state. In the light of the theory just discussed this property becomes comprehensible. When two atoms interact, the action of the perturbation can cause one of them to make a transition into an excited state. If the energy gained in forming a compound with an atom in an excited state is larger than that for a compound with the atom in the normal state, then the former will correspond to a more stable molecular configuration.

Thus, for example, we have seen in §73 that the stable term of the carbon atom is 3P_0 . Carbon atoms have two unpaired electrons in the $2p$ -state, and the valence in the normal state is equal to two. However, the carbon atom has an excited state with the configuration $1s^2 2s 2p^3$, in which the atom is in the state 5S . This state is higher than the normal state by 4.2 eV. In the 5S -state carbon has four unpaired electrons and its valence is equal to four. In compounds for which an energy larger than 4.2 eV is gained, carbon has valence four.

Another property of molecules following from the general theory is their geometric form. Formation of a chemical compound is related to the values of the Coulomb and exchange integrals which contain the products of the wave functions. If among these wave functions there are wave functions of electrons in the p -state, which have anisotropy in space (see §38), then the largest overlap of the wave functions is reached in selected spatial directions. As an example we can mention the molecule NH_3 . The nitrogen atom, having the configuration $1s^2 2s^2 2p^3 - ^4S$, has three p -electrons responsible for chemical binding. The wave functions of these electrons have their largest value in three mutually perpendicular directions. In the molecule NH_3 the angles between the bonds $N-H$ are close (although not exactly equal) to 90° .

The last fact important for understanding the structure of molecules is that the wave function of an excited state represents a linear combination of the wave functions of electrons. For example, in the excited state of the carbon atom mentioned above the wave function is a linear combination of the wave functions of one $2s$ -electron and three $2p$ -electrons. $2s$ - and $2p$ -states may be involved with different weights in forming the wave function.

This fact, called the hybridization of states, makes it possible, for instance, to understand why all four valences of the carbon atom are completely identical to each other.

Referring the reader to the specialist literature* for details, we stress that very great mathematical difficulties arise in attempting quantitative calculations of the forces of chemical binding and of the structure of the molecules formed. These difficulties are associated with the fact that the interaction responsible for the formation of a chemical bond cannot be considered as a small perturbation.

§80. The interaction of atoms at large distances

In addition to the forces of chemical character which in certain cases bind atoms into molecules, there is a weak interaction between atoms which are at relatively large distances from each other. Let us consider two widely separated atoms. Because of their spherical symmetry, the atoms have no mean dipole moments. However, the non-diagonal matrix elements of the dipole moments are different from zero.

One can imagine dipole moments of atoms arising in an obvious way as a result of the quantum-mechanical motion of the electrons, with the appearance and disappearance of a dipole moment, equal to zero only on the average. As a result instantaneous dipole moments are induced in both atoms. As an obvious illustration, one often considers the interaction of two oscillators in which the induced dipole moments are directly expressed in terms of their zero-point oscillations.

The interaction energy of atoms can be calculated by perturbation theory. As the perturbation operator one takes the energy of interaction of two dipoles given by formula (17.12) of Part I. The first order correction to the energy is equal to

$$E_1 = H'_{mm} = \int \psi^* \left\{ \frac{\mathbf{d}_1 \cdot \mathbf{d}_2 - 3(\mathbf{d}_1 \cdot \mathbf{n})(\mathbf{d}_2 \cdot \mathbf{n})}{R^3} \right\} \psi dV =$$

$$= \frac{\bar{\mathbf{d}}_1 \cdot \bar{\mathbf{d}}_2 - 3(\bar{\mathbf{d}}_1 \cdot \mathbf{n})(\bar{\mathbf{d}}_2 \cdot \mathbf{n})}{R^3} = 0$$

* See, for example, H.Eyring, I.Walter and G.Kimbal, *Quantum chemistry* (Wiley, New York, 1944, 1958); U.Kosman, *Introduction to quantum chemistry* (Academic Press, New York, 1957).

by virtue of $\bar{\mathbf{d}}_1 = \bar{\mathbf{d}}_2 = 0$. The second order correction can be written in the form

$$E_2 = \frac{1}{R^6} \sum_m \frac{[\mathbf{d}_1 \cdot \mathbf{d}_2 - 3(\mathbf{d}_1 \cdot \mathbf{n})(\mathbf{d}_2 \cdot \mathbf{n})]_{mn}^2}{E_n^{(0)} - E_m^{(0)}} = \frac{A}{R^6}.$$

According to the results of §53, the quantity

$$A = \sum_m \frac{[\mathbf{d}_1 \cdot \mathbf{d}_2 - 3(\mathbf{d}_1 \cdot \mathbf{n})(\mathbf{d}_2 \cdot \mathbf{n})]_{mn}^2}{E_n^{(0)} - E_m^{(0)}},$$

is always negative. Hence finally

$$U_{\text{vdW}} = E_2 = -\frac{\text{const}}{R^6}$$

This formula expresses the law of the van der Waals interaction. This interaction has no specific character, in the sense that it corresponds to attractive forces which decrease as R^{-7} for all atoms irrespective of their nature. The value of the constant can be expressed in terms of the polarizability of the atoms and varies for different atoms. Thus the van der Waals interaction between atoms represents the same specific quantum-mechanical effect as the chemical interaction. It cannot be understood on the basis of classical concepts, since atoms do not have a 'permanent' dipole moment.

The van der Waals forces, in contrast to forces leading to the formation of a chemical bond, possess additivity. If the interaction involves not two but three and more atoms, then the energy of interaction of the system, as any other weak perturbation, is obtained by the addition of the energies of pair interactions.

This result is of a general character, since we have not used a particular wave function.

However, if the atoms are not in S-states, then they can have a mean quadrupole moment different from zero. In this case, in addition to the van der Waals interaction a quadrupole-quadrupole interaction $\sim R^{-5}$ will exist between the atoms.

As distinct from atoms, molecules may have a mean dipole moment. If, however, its value as well as the value of the quadrupole moment is small, then the formula for U_{vdW} also applies to the molecules.

A particular situation arises when two identical atoms in different states interact, say, when an excited and a non-excited atom of one and the same

element interact. In this case an additional degeneracy, associated with the possibility of excitation exchange between the atoms, arises in the system.

The perturbation operator in this case is also the dipole-dipole interaction operator. However, the interaction energy is defined not by the mean value of this operator but by the solution of the corresponding secular equation (see §54). If the given atoms have non-zero matrix elements of the transition between the ground state and the excited state considered, then the interaction energy already turns out to be different from zero in the first approximation of perturbation theory. In this case the dipole-dipole interaction with resonant transfer of excitation takes place between the atoms. As can easily be seen, the energy of this interaction decreases only in inverse proportion to the cube of the distance between the atoms, $U \sim R^{-3}$.

Suppose, for example, that one of the atoms is in the ground 1S_0 state, while the other is in the excited 1P_1 state. We denote the wave functions corresponding to these states respectively by φ and ψ_m (the subscript m characterizes the angular momentum component in the state 1P_1 , $m = -1, 0, 1$). Thus a system of two non-interacting atoms turns out to be six-fold degenerate. It is described by unperturbed functions of the form $\varphi(1)\psi_m(2)$ and $\varphi(2)\psi_m(1)$. The matrix elements of the interaction operator \hat{H}'

$$\hat{H}' = \frac{\mathbf{d}_1 \cdot \mathbf{d}_2}{R^3} - \frac{3(\mathbf{d}_1 \cdot \mathbf{R})(\mathbf{d}_2 \cdot \mathbf{R})}{R^5}$$

with respect to these wave functions are not equal to zero for transitions between states differing by excitation transference. The calculations are conveniently carried out in a system of coordinates with the z -axis direct along the vector \mathbf{R} . Solving a secular equation of the form (54.4), we find the expression for the interaction energy

$$U_1 = \pm \frac{g^2}{R^3}, \quad U_2 = \pm \frac{2g^2}{R^3}, \quad (80.1)$$

where g denotes a matrix element of the form

$$g = \int \psi_0^* d_z \varphi dV.$$

The upper signs in formulae (80.1) refer to symmetric, and the lower signs to antisymmetric, excitation states. Energies U_1 correspond to states with $\Lambda = 1$, while energies U_2 correspond to states with $\Lambda = 0$.

In gaseous systems in which there is a considerable concentration of excited atoms the dipole-dipole interaction with resonant transfer of excitation can play a more important role than the van der Waals interaction. It

does not disappear in correct averaging over the orientations of the dipole moment of the atom and gives a basic contribution to the thermodynamic functions of the system*. The resonant dipole-dipole interaction is not additive.

§81. The comparison of molecular terms with atomic terms

The states of a molecule formed from two atoms can be related to the states of the atoms if the process of formation of the molecule is imagined as a result of their infinitely slow approach to each other.

The angular momentum component along the axis joining the two nuclei is conserved in the course of the process. On the other hand, as we have seen before, the component of the total angular momentum Λ along this axis will also be conserved for this molecule (see §78). We shall determine the possible values of Λ , as well as the number of energy states of the molecule formed.

Let the atoms be characterized by total angular momenta L_1 and L_2 , respectively. We assume that $L_1 > L_2$. The components of the angular momenta of the atoms can take on, respectively, the following values:

$$M_1 = L_1, \quad L_1 - 1, \quad L_1 - 2, \quad \dots, \quad -L_1,$$

$$M_2 = L_2, \quad L_2 - 1, \quad L_2 - 2, \quad \dots, \quad -L_2,$$

In accordance with the definition of the quantity Λ (see §78), there corresponds to the maximum value $\Lambda = L_1 + L_2$ the only state in which the components of the angular momenta of the atoms are equal to $M_1 = L_1$, $M_2 = L_2$. The next possible value of Λ is equal to $\Lambda = L_1 + L_2 - 1$. To this value of Λ there correspond two terms arising respectively from two states; in the first $M_1 = L_1$, $M_2 = L_2 - 1$, and the second $M_1 = L_1 - 1$, $M_2 = L_2$. Analogously, to the value $\Lambda = L_1 + L_2 - 2$ there correspond 3 terms arising from the states: $M_1 = L_1$, $M_2 = L_2 - 2$; $M_1 = L_1 - 1$, $M_2 = L_2 - 1$; $M_1 = L_1 - 2$, $M_2 = L_2$. The results obtained are conveniently expressed in the table:

for	$\Lambda = L_1 + L_2$	1 term is possible ,
for	$\Lambda = L_1 + L_2 - 1$	2 terms are possible ,
for	$\Lambda = L_1 + L_2 - 2$	3 terms are possible ,
⋮		
for	$\Lambda = L_1 - L_2$	$2L_2 + 1$ terms are possible .

* V.I.Malnev and S.I.Pekar, Soviet Physics JETP 24 (1967) 1220; 31 (1970) 597; Yu.A.Vdovin, Soviet Physics JETP 27 (1968) 242.

We can see by a simple calculation that the number of terms for $\Lambda < L_1 - L_2$ is equal to $2L_2 + 1$ and does not depend on Λ . In determining all possible states of the system account must be taken of the fact that each energy level with $\Lambda \neq 0$ is degenerate, since the energy of the system cannot depend on the orientation of the angular momentum in space.

The Σ term requires particular consideration.

A molecule turns out to be in the Σ state if $M_1 = -M_2$. This condition is fulfilled in L_2 cases where we have for the angular momentum components $M_1 > 0$ and $M_2 < 0$ and also in L_2 cases where $M_1 < 0$ and $M_2 > 0$. Furthermore, M_1 and M_2 can be equal to zero. Consequently, in the Σ state the molecule can also be formed from $2L_2 + 1$ energy states.

In §78 we have pointed out that Σ terms are divided into Σ^+ and Σ^- terms, depending on the symmetry properties of the system. The symmetry properties of the system do not change when the atoms are put an infinite distance apart. Hence the wave functions of the system for the states $|M_1| = |M_2|$ can be written in the form of symmetric or antisymmetric combinations

$$\psi_s = \psi_M^{(1)} \psi_{-M}^{(2)} + \psi_{-M}^{(1)} \psi_M^{(2)}, \quad (81.1)$$

$$\psi_a = \psi_M^{(1)} \psi_{-M}^{(2)} - \psi_{-M}^{(1)} \psi_M^{(2)}. \quad (81.2)$$

The Σ state corresponding to the values $M_1 = M_2 = 0$ is determined by the behaviour of the function $\psi = \psi_0^{(1)} \psi_0^{(2)}$ under reflection in a plane joining the nuclei of the atoms. Depending on the actual properties of the wave functions $\psi_0^{(1)}$ and $\psi_0^{(2)}$ there arise Σ^+ or Σ^- terms. Thus in L_2 cases a molecule in the Σ^+ state is formed, while in another L_2 cases a molecule in the Σ^- state is formed. One more Σ^+ or Σ^- term arises depending on the form of the function $\psi_0^{(1)} \psi_0^{(2)}$.

So far we have considered molecules formed from two different atoms. If the molecule is made up of identical atoms, then the calculation of its possible states is somewhat modified. Two cases are possible; either the separated atoms are in different states or they are in identical states. In the first case the number of possible terms must be doubled in comparison with the number of terms of a molecule consisting of different atoms, since the state of a molecule made up of identical atoms is invariant under the inversion transformation, and even and odd terms can be formed. If the atoms are in identical states, then the total number of states remains the same as for a molecule with different atoms. The problem of the parity of these states is rather complex*.

* See L.D. Landau and E.M. Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965) or E. Wigner and E. Witmer, *Z.f. Phys.* 51 (1958) 859.

§ 82. Rotation and vibration of diatomic molecules

We can now turn to the quantitative consideration of the motion of the nuclei in diatomic molecules. We shall not be interested in the translational motion of the molecule as a whole.

The motion of the nuclei in a molecule depends only on the distance between the nuclei. In the adiabatic approximation, according to eq. (78.3), the wave function satisfies the Schrödinger equation, which in spherical coordinates has the form

$$\left[-\frac{\hbar^2}{2\mu} \frac{1}{R^2} \frac{\partial}{\partial R} \left(R^2 \frac{\partial}{\partial R} \right) + E_n(R) + \frac{\hbar^2 \hat{K}^2}{2\mu R^2} \right] \alpha_n = 0, \quad (82.1)$$

where $E_n(R)$ is the electron energy, and \hat{K} is the angular momentum operator of the motion of the nuclei. We assume the electron energy to be fixed and consider the nuclei for a given E_n . Then the motion of the nuclei amounts to rotations and vibrations about an equilibrium position. The angular momentum operator of the nuclei must be expressed in terms of the total angular momentum operator of the molecule, which can be written in the form $\hat{\mathbf{J}} = \hat{\mathbf{K}} + \hat{\mathbf{L}}$, where $\hat{\mathbf{L}}$ is the angular momentum of the system of electrons.

It is clear that the molecule is in a state with a definite value of the total angular momentum. The angular momentum of the nuclei can run over a sequence of values corresponding to different rotational states of the electrons. Hence we shall be interested only in the mean value of the quantity

$$\overline{(\hat{K})^2} = \overline{(\hat{\mathbf{J}} - \hat{\mathbf{L}})^2} = J^2 + \overline{(\hat{L})^2} - 2\overline{\hat{\mathbf{J}} \cdot \hat{\mathbf{L}}}$$

since \hat{J}^2 has a definite value and is conserved.

According to what was said in § 78, the component of the angular momentum of the electrons along the axis of the molecule, $L_z = \Lambda$, is conserved.

The two other components are on the average equal to zero: $\overline{L_x} = 0$, $\overline{L_y} = 0$. Hence we have $\overline{L^2} = \Lambda^2$. Furthermore, since the direction of the vector \mathbf{n} (the axis of the molecule) is the only specified direction, the following equality holds:

$$\overline{\hat{\mathbf{L}}} = \mathbf{n} \Lambda.$$

The vector of the mean angular momentum of the nuclei in a diatomic molecule is perpendicular to \mathbf{n} , i.e.

$$\hat{\mathbf{K}} \cdot \mathbf{n} = (\hat{\mathbf{J}} - \hat{\mathbf{L}}) \cdot \mathbf{n} = 0.$$

(This statement follows directly from the fact, known from classical mecha-

tics, that the angular momentum vector in a two-body system is perpendicular to the axis joining to two bodies.) Hence it follows that

$$\hat{\mathbf{J}} \cdot \mathbf{n} = \hat{\mathbf{L}} \cdot \mathbf{n} = \Lambda. \quad (82.2)$$

From (82.2) we obtain

$$(\hat{\mathbf{J}} \cdot \hat{\mathbf{L}}) = \Lambda^2.$$

Finally, we find

$$(\hat{K})^2 = \hat{J}^2 + \hat{L}^2 - 2\Lambda^2 = J(J+1) + \bar{L}^2 - 2\Lambda^2. \quad (82.3)$$

where the quantum number J runs over a sequence of integers $J \geq \Lambda$.

In formula (82.3) the last two terms depend only on the state of the system of electrons, whereas the first term characterizes the rotation of the molecule as a whole.

The Schrödinger equation assumes the form

$$\left\{ -\frac{\hbar^2}{2\mu} \frac{1}{R^2} \frac{\partial}{\partial R} \left(R^2 \frac{\partial}{\partial R} \right) + E_n(R) + \frac{\hbar^2(\bar{L}^2 - 2\Lambda^2)}{2\mu R^2} + \frac{\hbar^2 J(J+1)}{2\mu R^2} \right\} \alpha_n = 0. \quad (82.4)$$

Denoting

$$U(R) = E_n(R) + \frac{\hbar^2(\bar{L}^2 - 2\Lambda^2)}{2\mu R^2} \quad (82.5)$$

we see that $U(R)$ plays the role of an effective potential energy. We shall consider those states of the nuclei for which the distance between the nuclei remains close to the equilibrium distance.

The effective potential energy can be written in the form

$$U(R) \cong U(R_0) + \frac{d^2 U}{dR^2} \Big|_{R=R_0} \frac{1}{2} (R - R_0)^2 = U(R_0) + \frac{1}{2} \mu \omega_0^2 (R - R_0)^2,$$

where ω_0 is the frequency of vibration. Eq. (82.4) finally assumes the form

$$\left[-\frac{\hbar^2}{2\mu} \frac{\partial}{\partial R} \left(R^2 \frac{\partial}{\partial R} \right) + U(R_0) + \frac{1}{2} \mu \omega_0^2 (R - R_0)^2 + \frac{\hbar^2}{2\mu R_0^2} J(J+1) \right] \alpha = 0. \quad (82.6)$$

We see that in the adiabatic approximation used, the motion of the molecule for a given electron state amounts to its rotation as a whole together with harmonic vibrations.

The total energy of the molecule is given by the formula

$$E = E^{\text{el}} + \frac{\hbar^2}{2\mu R_0^2} J(J+1) + \hbar \omega_0 (v + \frac{1}{2}), \quad (82.7)$$

where v is the vibrational quantum number.

Let us also estimate the degree of accuracy of the adiabatic approximation. According to (57.6) the parameter of non-adiabaticity is in order of magnitude

$$\hat{C}\alpha \sim \frac{\hbar^2}{\mu} \nabla\alpha \int \varphi^* \nabla\varphi dV. \quad (82.8)$$

We are interested in the dependence of this expression on the reduced mass μ . We estimate in order of magnitude the derivative $\nabla\alpha$ in the vibrational ground state. Evidently, $\nabla\alpha \sim \alpha[(R-R_0)^2]^{-\frac{1}{2}}$. In order to estimate the mean displacement $[(R-R_0)^2]^{\frac{1}{2}}$ we note that in the ground state the mean potential energy is equal to one half of the total energy, i.e.

$$\frac{1}{2}\mu\omega_0^2 \overline{(R-R_0)^2} = \frac{1}{2}\hbar\omega_0.$$

Hence $[(R-R_0)^2]^{\frac{1}{2}} \sim (\omega_0\mu)^{-\frac{1}{2}}$. Thus $\nabla\alpha \sim (\omega_0\mu)^{\frac{1}{2}}\alpha$. But by definition

$$\omega_0 = [\mu^{-1}(d^2U/dR^2)_{R=R_0}]^{\frac{1}{2}} \sim \mu^{-\frac{1}{2}}.$$

The integral in (82.8) depends only on the electron part of the system and does not depend on μ . Hence, finally, $C \sim \mu^{-\frac{3}{4}}$. From dimensionality considerations it follows that

$$C \sim (m/\mu)^{\frac{3}{4}}. \quad (82.9)$$

Indeed, the total Schrödinger equation describing the motion of all particles in the molecule involves only two quantities of the dimensionality of mass; μ and the electron mass m . One cannot construct any other quantities of the dimensionality of mass from the quantities involved in the Schrödinger equation. Thus the parameter C is very small even for the hydrogen molecule. The quantity $(m/\mu)^{\frac{3}{4}}$ is the basic small parameter of the theory of molecules.

The spacings between electron energy levels ΔE^{el} for molecules do not depend on the mass of the nuclei and are of the same order of magnitude as for atoms (i.e. are of the order of a few eV).

The spacing between vibrational levels is

$$\Delta E^{\text{vib}} = \hbar\omega_0 \sim \mu^{-\frac{1}{2}} \ll \Delta E^{\text{el}}. \quad (82.10)$$

This amounts to a few tenths of an electronvolt.

Finally, the spacing between rotational levels is

$$\Delta E^{\text{rot}} = \frac{\hbar^2}{2\mu R_0^2} [J'(J'+1) - J(J+1)] \quad (82.11)$$

where $\Delta J = J' - J = \pm 1$.

Since $\Delta E^{\text{rot}} \sim \mu^{-1}$, this spacing is much smaller than that between vibrational levels and amounts to few meV.

Knowing the distribution of levels, one can find the emission (or absorption) spectra of molecules, which differ strongly in character from atomic line spectra. However, one then has to take into account an important fact. In ch. 12, devoted to radiation theory, it will be shown that transitions between levels are limited by so-called selection rules. It turns out that transitions are possible only between levels for which a change in the quantum numbers defined by the conditions

$$\begin{array}{ccc} & J+1 \\ & \nearrow \\ J' & \rightarrow J & v' = v \pm 1 \\ & \searrow \\ & J-1 \end{array}$$

takes place (except for the transition $J' = J = 0$, which is forbidden).

Taking into account the selection rules, the proof of which will be given in ch. 12, it is possible to find the frequencies emitted or absorbed. In the case of transitions for which the electron state of the molecule does not change, we obtain (82.7)

$$\hbar\omega = \hbar\omega_n(v'-v'') + B[J'(J'+1) - J''(J''+1)], \quad (82.12)$$

where $B = \hbar/2\mu R^2$. Taking into account the selection rules, we obtain two frequency branches for a given difference $v' - v''$. For $J'' = J' + 1$ we have the first branch of frequencies:

$$\hbar\omega_1 = \hbar\omega_n(v'-v'') - 2B(J'+1), \quad J' = 0, 1, 2, \dots \quad (82.13)$$

For $J'' = J' - 1$ we find the second branch of frequencies:

$$\hbar\omega_2 = \hbar\omega_n(v'-v'') + 2BJ', \quad J' = 1, 2, 3, 4, \dots \quad (82.14)$$

We note that J' cannot be equal to zero, since this would correspond to $J'' = -1$.

Let us consider the order and distribution of these frequencies for a given difference $v' - v''$. The frequency ω_1 decreases beginning with $\omega_1 = \omega_n(v' - v'') - 2B$, and the frequency ω_2 increases from the lowest value equal to $\omega_n(v' - v'') + 2B$. The spacing between the lines in each branch is equal to $2B$, while the spacing between the branches is equal to $4B$. The frequency $\omega_n(v' - v'')$ lying between the bands is not observed. The set of lines ω_1 is also called the P-branch of frequencies, and the set of frequencies ω_2 is the R-branch. These frequencies lie in the infrared part of the spectrum.

We now turn to the study of frequencies which arise in transitions asso-

ciated with a change in the electron state. The character of such a spectrum differs fundamentally from the infrared spectrum considered. Radiation frequencies are defined in this case by the formula

$$\hbar\omega = E_0 + \hbar\omega_n(v' + \frac{1}{2}) - \hbar\omega_m(v'' + \frac{1}{2}) + B_n(J'+1)J' - B_mJ''(J''+1). \quad (82.15)$$

Here it should be stressed that $\omega_n \neq \omega_m$, and $B_n \neq B_m$. As a matter of fact, the vibration frequencies ω and the quantities B are determined by the electron state of the molecule and, consequently, when this state changes these quantities change substantially.

Since the change in the energy of the molecule in transitions associated with a change in the electron state is rather large, the frequencies observed in this case lie in the visible part of the spectrum. The set of lines corresponding to the chosen pair of quantum numbers v'_l and v''_l is called a band. The band is in its turn made up of three branches. These branches are obtained in the following way. In accordance with the selection rules the quantum number J'' can be equal to $J'' = J' - 1$, $J'' = J' + 1$ and $J'' = J'$. To the first case there corresponds the R-branch, whose frequencies are defined by the relation

$$\omega_1 = A + \bar{B}J'^2 + CJ', \quad (82.16)$$

where

$$A = E_0 + \hbar\omega_n(v' + \frac{1}{2}) - \hbar\omega_m(v'' + \frac{1}{2}),$$

$$\bar{B} = B_n - B_m, \quad C = B_n + B_m.$$

Transitions $J'' = J' + 1$ constitute the Q-branch, and frequencies are in this case defined by the formula

$$\hbar\omega_2 = A + \bar{B}J'(J'+1). \quad (82.17)$$

and, finally, for the P-branch we obtain

$$\hbar\omega_3 = A - 2B_m + \bar{B}J'^2 + (B_n - 3B_m)J' \quad (82.18)$$

In all three cases ω is a quadratic function of the quantum number J . To examine the distribution of frequencies it is convenient to refer to the diagram of fig. V.24. Here a parabola corresponding to eq. (82.18) is shown for $\bar{B} > 0$. The quantum number J is plotted on the vertical axis, and the frequency ω on the horizontal axis. Experimentally observed frequencies can easily be obtained by means of this diagram. If the points of intersection of horizontal lines drawn through integer values of J' and the parabola are projected on the horizontal axis, then we obtain the observed values of frequencies. These radiation frequencies are shown below the horizontal axis.

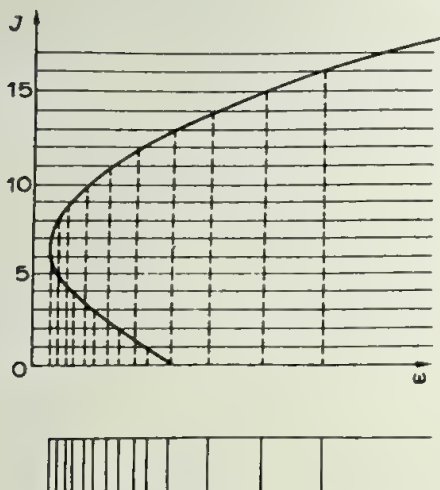


Fig. V.24

We see that the frequencies of the spectrum are not equally spaced, as was the case in the infrared part of the spectrum, but become denser in a certain part of it; as ω increases the spacings between the observed frequencies become larger. The place of convergence of the lines is called the band edge. In the case given the band edge lies on the low frequency side of the spectrum. If $B_n < B_m$, then the parabola is curved in the opposite way. In this case the band edge lies on the high frequency side. However, it should be stressed that many exceptions to the above rule are observed. For example, in the case where electron terms are degenerate more than three bands are observed; sometimes, for example, the Q-branch is not observed and so on*.

* For a detailed exposition of these problems see L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

Scattering Theory

§83. Scattering amplitude and cross section

We mean by a scattering process the deflection of particles from their initial directions of motion caused by their interaction with a system which we shall call the scatterer.

The investigation of scattering processes of charged and neutral particles is one of the basic experimental methods of studying the structure of atoms, atomic nuclei and elementary particles.

Indeed, the very existence of the atomic nucleus was established in Rutherford's experiments on the scattering of α -particles. The analysis of the results of experiments on the scattering of neutrons by nuclei enabled Bohr to formulate modern concepts on the structure of the nucleus. The study of the laws of scattering of fast particles is the main source of information about nuclear forces and of the properties of elementary particles.

From these examples, although they are far from being complete, it is easy to evaluate the significance of scattering theory, one of the most important branches of quantum mechanics.

The scattering of a flux of particles is characterized by the differential scattering cross section. This quantity is defined as the ratio of the number of particles dN_{scat} scattered per unit time into the solid angle $d\Omega$ to the flux density of the incident particles, j_{inc} , i.e. the differential cross section is defined by the relation

$$d\sigma(\theta, \varphi) = \frac{dN_{\text{scat}}(\theta, \varphi)}{j_{\text{inc}}},$$

where the angles θ and φ define the direction of motion of the scattered particles. The z -axis is taken along the direction of motion of the incident particles.

For our purposes it is convenient to write dN_{scat} in the form

$$dN_{\text{scat}}(\theta, \varphi) = j_{\text{scat}}(\theta, \varphi) ds,$$

where j_{scat} is the flux density of the scattered particles at large distances from the scattering centre, and ds is an element of area perpendicular to the radius vector drawn from the scattering centre at the angles θ, φ . The quantity ds is related to the solid angle element $d\Omega$ by the equality

$$ds = r^2 d\Omega.$$

Thus the differential cross section is defined by the formula

$$d\sigma = \frac{j_{\text{scat}}}{j_{\text{inc}}} ds. \quad (83.1)$$

This definition of the cross section is the same as that introduced in §43 of Part I. In quantum mechanics j_{scat} and j_{inc} are probability current densities defined in §7.

In the mutual scattering of two quantum-mechanical systems, for example the scattering of an electron by an atom, a neutron by a nucleus, an atom by an atom and so on, one has to distinguish between elastic and inelastic scattering. In elastic scattering the internal state of both the scatterer and the scattered system remains unchanged. For example, in the elastic scattering of electrons by atoms the state of the latter remains unchanged. In inelastic scattering the internal state of one or both systems changes. For example, the scattering of electrons by atoms is inelastic if in the process of scattering the atoms make a transition into an excited state.

In inelastic scattering a part of the kinetic energy goes over into the internal energy or, conversely, the internal energy goes over into kinetic energy. Collisions of this latter type are called collisions of the second kind.

We shall begin the exposition of scattering theory with the simpler case of elastic scattering. In this case one need not be concerned with the internal state of the systems and can call the interacting systems particles (although these systems may have a complex internal structure, for example they may be atoms, molecules or nuclei).

In the scattering process there is an interaction between two particles; the scattered particle and the scatterer. In this case the potential energy of the

interaction very often has the form $U(r)$. Then the problem of the motion of two interacting particles can always be reduced to the study of the motion of one particle (with reduced mass μ) in the field of the motionless centre of force at the centre of mass of the system.

In practice it is always necessary to know how the process appears in the laboratory system of coordinates. Therefore if the problem of motion of one particle in the field of external forces is solved, then in the final formulae one has to transform to the laboratory system. This can easily be done, knowing that the cross section (83.1) is invariant under transformation from one Galilean system to another, and that the angles transform by means of the relations (see §43 of Part I)

$$\tan \theta_1 = \frac{m_1 \sin \theta}{m_1 + m_2 \cos \theta}; \quad \theta_2 = \frac{\pi - \theta}{2}. \quad (83.2)$$

Here θ is the scattering angle of the two particles in the centre-of-mass system; θ_1 and θ_2 are the scattering angles of the first and second particles in the laboratory system of coordinates, in which the second particle was at rest before the collision.

We now consider the wave function of the particle scattered by the force centre. We shall not, as yet, make any assumptions about the concrete form of the potential energy of interaction.

We let the motionless scattering centre be at the origin and take the direction of the incident particle flux to be the z -axis. At a distance from the scattering centre the incident particle moves as a free particle, and its wave function has the form of a plane wave e^{ikz} . Near the force centre the particle undergoes scattering and the form of its wave function is different.

However, when the scattered particle goes sufficiently far from the force centre, it will again move as a free particle. Since the scattered particle flux at a large distance will always be directed from the scattering centre, the motion of the scattered particles must be described by a diverging wave $f(\theta, \varphi) e^{ikr}/r$.

The total wave function describing the motion of the incident and scattered particles at large distances from the scattering centre can be written in the form

$$\psi = e^{ikz} + f(\theta, \varphi) \frac{e^{ikr}}{r}, \quad (83.3)$$

where the first term describes the motion of the incident particles, and the second term the motion of the scattered particles.

The amplitude of the diverging wave $f(\theta, \varphi)$, called the scattering amplitude,

depends, generally speaking, on the angles θ and φ . According to (83.1), one has to calculate the incident and scattered particle flux densities. In accord with formula (7.6), the flux density in the plane wave e^{ikz} incident on the scattering centre is equal to $p/m = v$, where v is the velocity of the particles. Analogously, the flux density in the diverging wave is given by the expression

$$\frac{|f(\theta, \varphi)|^2 v}{r^2}. \quad (83.4)$$

Taking the ratio of the incident and scattered fluxes we obtain, in correspondence with formula (83.1), the differential cross section

$$d\sigma = |f(\theta, \varphi)|^2 d\Omega. \quad (83.5)$$

Thus we see that the cross section is completely determined by the value of the scattering amplitude. The calculation of the latter is usually carried out in the following way. One finds the solution of the Schrödinger equation for the motion of the particle in the field of the scattering centre, which at large distances from the centre has the form (83.3). Then the coefficient of the factor e^{ikr}/r gives the scattering amplitude to be determined.

The wave function describing the motion of the particle at a distance from the scattering centre was written in the form (83.3), i.e. as the sum of incident and diverging waves, on the basis of simple and obvious physical considerations.

However, it can be shown rigorously that at a large distance from the fixed scattering centre $U(r)$ the solution of the Schrödinger equation has indeed the form (83.3). For this we write the Schrödinger equation in the form

$$(\nabla^2 + k^2)\psi = \frac{2mU}{\hbar^2} \psi,$$

where $k^2 = 2mE/\hbar^2$, and m and E are respectively the mass and energy of the scattered particle. By means of a Green's function the solution can (see §15) be written in the form

$$\psi = \psi_0 + \int G(\mathbf{r}, \mathbf{r}') \frac{2m}{\hbar^2} U(\mathbf{r}') \psi(\mathbf{r}') dV', \quad (83.6)$$

where the function ψ_0 satisfies the equation

$$(\nabla^2 + k^2)\psi_0 = 0.$$

The solution of this equation evidently has the form of a plane wave e^{ikz} .

The Green's function satisfies the equation

$$(\nabla^2 + k^2)G(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}').$$

This equation is formally identical with eq. (24.20) of Part I, if ω^2/c^2 in it is replaced by k^2 and $-4\pi j_0/c$ by $\delta(\mathbf{r} - \mathbf{r}')$. Without reproducing the calculations of § 24 of Part I, we make use of a formula similar to eq. (24.19) of Part I and write the solution for $G(\mathbf{r}, \mathbf{r}')$ in the form

$$G(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi} \int \frac{\delta(\mathbf{r}'' - \mathbf{r}') e^{ik|\mathbf{r} - \mathbf{r}''|} dV''}{|\mathbf{r} - \mathbf{r}''|}.$$

Carrying out the integration with respect to dV'' , we obtain

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi} \frac{e^{ik|\mathbf{r} - \mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|}.$$

Substituting the values of ψ_0 and G into (83.6), we arrive at the integral equation

$$\psi = e^{ikz} - \frac{m}{2\pi\hbar^2} \int \frac{e^{ik|\mathbf{r} - \mathbf{r}'|} U(\mathbf{r}') \psi(\mathbf{r}') dV'}{|\mathbf{r} - \mathbf{r}'|}. \quad (83.7)$$

Further, we consider the integral involved in formula (83.7) and determine its values at large distances \mathbf{r} . We define large distances in the following way. Let the range of values of \mathbf{r}' in which the integrand differs considerably from zero and which gives the basic contribution to the value of the integral be R . We shall call large distances those distances $|\mathbf{r}|$ for which the inequality

$$|\mathbf{r}| \gg R \quad (83.8)$$

is fulfilled; such distances always exist when $U(r)$ decreases sufficiently rapidly. In calculating integral (83.7) at large distances it can be assumed that $|\mathbf{r}| \gg |\mathbf{r}'|$.

Expanding $|\mathbf{r} - \mathbf{r}'|$ in a series, we have

$$|\mathbf{r} - \mathbf{r}'| = [(\mathbf{r} - \mathbf{r}')^2]^{\frac{1}{2}} = (r^2 - 2\mathbf{r} \cdot \mathbf{r}')^{\frac{1}{2}} = r - \frac{\mathbf{r} \cdot \mathbf{r}'}{r}.$$

Substituting this expansion into (83.7), we find

$$\psi = e^{ikz} - \frac{me^{ikr}}{2\pi\hbar^2 r} \int U(\mathbf{r}') e^{-ik \cdot \mathbf{r}'} \psi(\mathbf{r}') dV'. \quad (83.9)$$

Here $\mathbf{k} = kr/r$. The wave vector \mathbf{k} is evidently directed along the radius vector. It characterizes the direction of propagation of the diverging spherical wave. Comparing (83.9) with (83.3) we see that this last expression is of a general character. The scattering amplitude is equal to

$$f(\theta, \varphi) = -\frac{m}{2\pi\hbar^2} \int U(\mathbf{r}') \psi(\mathbf{r}') e^{-i\mathbf{k} \cdot \mathbf{r}'} dV'. \quad (83.10)$$

We shall need formulae (83.9) and (83.10) for what follows.

§84. The Born approximation

Although we could find the asymptotic expression for the wave function, the problem of obtaining the concrete form of the scattering amplitude is still far from being solved. Indeed, according to formula (83.10) the scattering amplitude is expressed in terms of the unknown wave function ψ . The exact solution of the Schrödinger equation and the determination of $f(\theta, \varphi)$ in most problems of practical interest is associated with very great mathematical difficulties. Therefore approximate methods are widely used in scattering theory. The most important of these is the Born approximation. This method is based on the assumption that the potential energy of interaction of the scattered particle with the force centre is small, so that it can be considered as a small perturbation.

If the potential energy is a small perturbation, then it can be assumed that the initial motion of the particle is changed only slightly. Then the integral equation (83.9) can easily be solved by a method of successive approximation. In the zero order approximation the small term containing the potential energy can be dropped. Then

$$\psi_0 = e^{ikz} = e^{i\mathbf{k}_0 \cdot \mathbf{r}}, \quad (84.1)$$

where \mathbf{k}_0 is a vector equal to $\mathbf{k}_0 = k\mathbf{n}_0$, and \mathbf{n}_0 is the unit vector along the z -axis. In the first approximation, in place of the wave function on the right-hand side of (83.9), one has to substitute the value of its zero order approximation (84.1). We obtain

$$\psi = e^{ikz} - \frac{m e^{ikr}}{2\pi\hbar^2 r} \int U(\mathbf{r}') e^{ikz'} - i\mathbf{k} \cdot \mathbf{r}' dV'. \quad (84.2)$$

In this approximation the scattering amplitude is equal to

$$f(\theta, \varphi) = -\frac{m}{2\pi\hbar^2} \int U(\mathbf{r}') e^{i\mathbf{K} \cdot \mathbf{r}'} dV'. \quad (84.3)$$

Here we have introduced the notation

$$\mathbf{K} = \mathbf{k}_0 - \mathbf{k}. \quad (84.4)$$

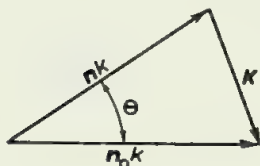


Fig. V.25

where, in correspondence with fig. V.25, the modulus of the vector \mathbf{K} is defined by the relation

$$K = k|\mathbf{n} - \mathbf{n}_0| = 2k \sin \frac{1}{2}\theta = \frac{2mv}{\hbar} \sin \frac{1}{2}\theta. \quad (84.5)$$

The vector \mathbf{K} is often called the collision vector. Correspondingly the vector $\mathbf{P} = \hbar\mathbf{K}$ is called the momentum transfer vector. If the potential energy does not depend on angles, $U = U(|\mathbf{r}|)$, then in (84.3) one can carry out the integration over angles

$$\begin{aligned} f(\theta) &= -\frac{m}{2\pi\hbar^2} \int_0^\infty U(|\mathbf{r}'|) r'^2 dr' \int_0^\pi e^{iKr' \cos \theta} \sin \theta d\theta \int_0^{2\pi} d\varphi = \\ &= -\frac{2m}{\hbar^2} \int_0^\infty U(|\mathbf{r}'|) \frac{\sin Kr'}{Kr'} r'^2 dr'. \end{aligned} \quad (84.6)$$

In the first approximation the scattering amplitude is determined by the potential energy to the first power. Then, if (84.3) is substituted into definition (83.5), we find

$$\begin{aligned} d\sigma &= |f(\theta)|^2 d\Omega = \frac{m^2}{4\pi^2\hbar^4} \left| \int_0^\infty U(|\mathbf{r}'|) e^{i\mathbf{K} \cdot \mathbf{r}'} dV' \right|^2 d\Omega = \\ &= \frac{4m^2}{\hbar^4} \left| \int_0^\infty U(|\mathbf{r}'|) \frac{\sin Kr'}{Kr'} r'^2 dr' \right|^2 d\Omega. \end{aligned} \quad (84.7)$$

Expression (84.7) is called the Born approximation. It is widely used in nuclear physics.

Continuing with successive approximations, i.e. substituting the wave function ψ from (84.2) into (83.9), it would be possible to find the wave function and scattering amplitude in the second approximation. The addition

to the scattering amplitude in the second approximation would be determined by the integral of the square of the potential energy of interaction. Corrections of higher orders can be found in an analogous way.

For small values of the scattering angle we have from (84.7)

$$d\sigma = \frac{4m^2}{\hbar^4} \left| \int_0^\infty U(|\mathbf{r}'|) r'^2 dr' \right|^2 d\Omega,$$

i.e. the cross section turns out to be independent of the velocity of the particle. In the next section we shall give an example of a concrete calculation of the cross section by the Born approximation. We now turn to a discussion of the applicability of the Born approximation.

For a rapid convergence of the series of successive approximations it is necessary that the correction to the wave function of the first approximation ψ_1 be small in comparison with the wave function of the zero order approximation ψ_0 , i.e. the following condition must be fulfilled:

$$|\psi_1| \ll |\psi_0|. \quad (84.8)$$

By means of (83.7) one can find the value of the function ψ_1 which is valid for arbitrary values of r ; then (84.8) is written in the form

$$\frac{m}{2\pi\hbar^2} \left| \int \frac{e^{ikz'} e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} U(\mathbf{r}') dV' \right| \ll 1. \quad (84.9)$$

Since $\psi_1(\mathbf{r}')$ decreases with distance from the scattering centre, then condition (84.9) will be fulfilled if it is fulfilled at the origin. Hence condition (84.9) can be replaced by the inequality

$$|\psi_1(0)| = \frac{m}{2\pi\hbar^2} \left| \int \frac{e^{ik(r'+z')} U(\mathbf{r}') dV'}{r'} \right| \ll 1. \quad (84.10)$$

Further estimates of the integral can be carried out in two limiting cases:

1. When the relation $kR \ll 1$ is satisfied, where R is the effective range of interaction. This corresponds to small energies of the particles

$$E \ll \frac{\hbar^2}{mR^2}.$$

2. When the inverse inequality $kR \gg 1$ is fulfilled. This corresponds to the condition

$$E \gg \frac{\hbar^2}{mR^2}.$$

In the first case, in estimating the integral one can set $e^{ik(r'+z')}$ in (84.10) equal to unity.

Then (84.10) gives in order of magnitude

$$\frac{m}{\hbar^2} \int \frac{|U(r')| dV'}{r'} \approx \frac{m}{\hbar^2} |U_0| R^2 \ll 1.$$

Here U_0 is a mean value of the interaction energy in the range R .

We write this last relation in the form

$$U_0 \ll \frac{\hbar^2}{mR^2}. \quad (84.11)$$

According to (37.9), the expression \hbar^2/mR^2 is equal in order of magnitude to the minimum depth of the potential well of radius R for which the level arises. We see that the condition of applicability of the Born approximation for the scattering of slow particles has a simple meaning. Namely, the mean interaction energy must be small in comparison with the minimum potential energy of the particle in the well for which the bound state is formed.

In the case of a large energy of the particle the region of applicability of the Born approximation is considerably enlarged. The exponential factor in formula (84.10) oscillates very rapidly, leading to a decrease in the total value of the integral.

In calculating the integral one can take the slowly varying factors out, writing

$$\begin{aligned} |\psi_1(0)| &= \frac{m}{2\pi\hbar^2} |U_0| \left| \iint \frac{e^{ikz'} e^{ikr'}}{r'} dV' \right| = \\ &= \frac{m|U_0|}{\hbar^2} \left| \int_0^R \int_0^\pi e^{ikr'(1+\cos\vartheta)} \sin\vartheta d\vartheta r' dr' \right| = \\ &= \frac{m|U_0|}{\hbar^2 k} \left| \int_0^R (1 - e^{2ikr'}) dr' \right| \approx \frac{m|U_0|R}{\hbar^2 k} \ll 1. \end{aligned}$$

Here we have dropped the integral of the rapidly oscillating quantity $e^{2ikr'}$ since it is small in comparison with the integral retained.

Rewriting this last inequality in the form

$$\frac{|U_0|R}{\hbar v} \ll 1, \quad (84.12)$$

we see that the Born approximation becomes valid for particles with larger

energy the larger the product $U_0 R$ determined by the properties of the scattering centre.

In the important case of the Coulomb field the potential Ze^2/r decreases so slowly that the concept of an effective range of interaction R cannot be introduced.

However, we note that for $U_0 = Ze^2/R$ the product $U_0 R$ contained in the inequality (84.12) does not depend on R .

Hence for the Coulomb field inequality (84.12) assumes the form

$$\frac{Ze^2}{\hbar v} \ll 1. \quad (84.13)$$

This has an obvious meaning: if the velocity of the electron in the first Bohr orbit of a hydrogen-like atom with nuclear charge Ze (the quantity $v_k = Ze^2/\hbar$) is introduced, then formula (84.13) assumes the form

$$\frac{v_k}{v} \ll 1, \quad (84.14)$$

i.e. the velocity of the particle must be large in comparison with the velocity of the electron in the first Bohr orbit.

For the Born approximation to be applicable inequality (84.13) requires a larger energy the larger the charge of the scattering nucleus.

§85. The scattering of fast charged particles by atoms

Let us apply the Born approximation to the calculation of the cross section for the scattering of fast charged particles by atoms.

We shall assume that the nucleus of the atom with charge Ze is at the origin, and that the charge of the atomic shell is distributed in space with density $n(\mathbf{r})$. We shall disregard the size of the nucleus treating it as a point.

The differential scattering cross section is given by formula (84.6), which for $U = e\varphi$, where φ is the potential of the electric field acting on the particle to be scattered and e is its charge, assumes the form

$$d\sigma = \frac{m^2 e^2}{4\pi^2 \hbar^4} \left| \int \varphi(\mathbf{r}') e^{i\mathbf{K} \cdot \mathbf{r}'} dV' \right|^2 d\Omega. \quad (85.1)$$

The integral in formula (85.1) is conveniently expressed in terms of the charge density distribution in the atom.

For this we note that $\int \varphi(\mathbf{r}') e^{i\mathbf{K} \cdot \mathbf{r}'} dV'$ represents the Fourier component of the potential. It can be expressed in terms of the Fourier component of

the charge density analogously to formula (24.25) of Part I which relates the Fourier component of the current density to the Fourier component of the potential.

We then have

$$d\sigma = \frac{4m^2 e^2}{K^4 \hbar^4} \left| \int \rho(\mathbf{r}') e^{i\mathbf{K} \cdot \mathbf{r}'} dV' \right|^2 d\Omega. \quad (85.2)$$

The charge density in the atom can be written in the form

$$\rho(\mathbf{r}) = Ze\delta(\mathbf{r}) - en(\mathbf{r}). \quad (85.3)$$

For the differential cross section we finally obtain

$$\begin{aligned} d\sigma &= \frac{4m^2 e^4}{\hbar^4 K^4} \left| \int Z e^{i\mathbf{K} \cdot \mathbf{r}'} \delta(\mathbf{r}') dV' - \int n(\mathbf{r}') e^{i\mathbf{K} \cdot \mathbf{r}'} dV' \right|^2 d\Omega = \\ &= \frac{4m^2 e^4}{\hbar^4 K^4} |Z - F(K)|^2 d\Omega, \end{aligned} \quad (85.4)$$

where

$$F(K) = \int n(\mathbf{r}') e^{i\mathbf{K} \cdot \mathbf{r}'} dV'. \quad (85.5)$$

The quantity F is called the atomic form factor. Its value is determined by the electron charge density distribution.

Substituting into (85.5) the value of the collision vector \mathbf{K} according to (84.5), we rewrite the differential cross section in the form

$$d\sigma = \left(\frac{e^2}{2mv^2} \right)^2 |Z - F(K)|^2 \frac{d\Omega}{\sin^4 \frac{1}{2}\theta}. \quad (85.6)$$

Let us first consider a particular case of formula (85.6). If the scattering takes place on a point nucleus without an electron shell, $n = 0$, then, consequently, $F = 0$. We then obtain for the differential cross section

$$d\sigma = \left(\frac{Ze^2}{2mv^2} \right)^2 \frac{d\Omega}{\sin^4 \frac{1}{2}\theta}. \quad (85.7)$$

This is the well-known Rutherford formula, which is obtained in classical mechanics. The Rutherford formula in the case given was obtained by means of the approximate Born method. However, it is interesting to note that the same expression is obtained in an exact solution of the problem*. Since the

* See, for example, N.F.Mott and H.S.W.Massey, *The theory of atomic collisions* (Clarendon Press, Oxford, 1965).

scattering cross section in the exact solution does not contain Planck's constant \hbar , the results given by classical and quantum physics must naturally be the same.

The fact that the cross section becomes infinity for scattering at infinitely small angles is associated with the slow change of the Coulomb potential. Hence particles are scattered no matter how far away they pass from the scattering centre. However, as we shall see later, in practice the screening effect of the electron shell ensures a finite value of the scattering cross section.

Let us now consider the atomic form factor (85.5). The effective range of integration in it has a size of the order of the atomic size a . Outside this range $n(\mathbf{r})$ reduces to zero. Hence for small angles θ , for which $Ka \ll 1$, the exponent in integral (85.5) can be expanded in a series. We then have

$$Z - F(K) = Z - Z - i\mathbf{K} \int n(\mathbf{r}') \mathbf{r}' dV' + \frac{1}{2} \int n(\mathbf{r}') (\mathbf{K} \cdot \mathbf{r}')^2 dV'. \quad (85.8)$$

In formula (85.8) the first two terms mutually cancel, since the charge of the electron shell of the atom is equal to the charge of the nucleus. The third term represents the dipole moment of the atom, which, as we have seen (see §72), is equal to zero. In the last term, integrating over angles we obtain

$$Z - F = \frac{2\pi K^2}{3} \int_0^\infty n(r) r^4 dr.$$

The differential cross section in the limiting case $Ka \ll 1$ will have the form

$$d\sigma = \left(\frac{4\pi m e^2}{3\hbar^2} \right)^2 \left| \int n(r) r^4 dr \right|^2 d\Omega.$$

Thus owing to the screening by the charge of the electron shell the differential cross section for small scattering angles turns out to be a finite and constant (independent of angle) quantity. On the contrary, for large scattering angles, when the inverse inequality $Ka \gg 1$ is fulfilled, the exponential in integral (85.5) begins to oscillate rapidly and the form factor turns out to be small. Neglecting it in comparison with Z , we arrive at (85.7). The screening of the nuclear charge is not manifested for large scattering angles.

As an example let us calculate the form factor for the hydrogen atom. According to §38, the charge density in the hydrogen atom in the ground state is equal to

$$n(r) = |\psi(r)|^2 = \frac{1}{\pi a^3} e^{-2r/a}, \quad a = \frac{\hbar^2}{me^2}.$$

Consequently, the form factor is defined by the integral

$$F(K) = \frac{1}{\pi a^3} \int e^{-2r/a} e^{i\mathbf{K} \cdot \mathbf{r}} r^2 dr \sin \vartheta d\vartheta d\varphi. \quad (85.9)$$

Directing the z -axis along the vector \mathbf{K} , we have

$$F(K) = \frac{1}{\pi a^3} \int e^{-2r/a} e^{iKr \cos \vartheta} r^2 dr \sin \vartheta d\vartheta d\varphi.$$

Carrying out the integration, we finally find

$$F(K) = \frac{16}{(4+K^2a^2)^2}.$$

Then the differential cross section for the hydrogen atom can be written in the form

$$d\sigma = \left(\frac{e^2}{2mv^2} \right)^2 \left[1 - \frac{16}{(4+K^2a^2)^2} \right]^2 \frac{d\Omega}{\sin^4 \frac{1}{2}\theta}.$$

The total cross section is obtained by integration over all values of the scattering angle.

For other atoms of the periodic system of elements the charge density and the potential of interaction with the scattered particle can be calculated by means of Hartree or Thomas–Fermi approximate methods. After this the calculation of the form factor can be carried out in accordance with formula (85.5).

§86. Partial wave scattering theory

In the preceding sections we have considered one form of approximate scattering theory.

In addition to approximate theories it is possible to develop an exact scattering theory, which is often called partial wave theory.

The general scheme of partial wave scattering theory does not differ from that assumed in §83. We consider the motion of a particle in the field of a scattering centre. We assume that the scattering field is spherically symmetric and that at a distance from the centre the incident particle is described by the plane wave e^{ikz} and the scattered particle by a diverging spherical wave. Let

the general solution of the Schrödinger equation in the centrally symmetric field be found. At a distance from the scattering centre this solution must be written in the form (83.3), i.e. in the form of an incident plane wave and a diverging spherical wave. As we know, the amplitude of the latter determines the scattering cross section which is of interest to us.

According to (35.31), the general solution of the Schrödinger equation in a centrally symmetric field independent of the angle φ can be represented by the expansion

$$\psi = \sum_{l=0}^{\infty} A_l R_l(r) P_l(\cos \theta). \quad (86.1)$$

We shall call a given term of the series (86.1) the l th partial wave. At a large distance from the force centre the asymptotic form of the radial functions R_l is given by formulae (35.26) and (35.27)

$$\begin{aligned} R_l &= B_l \frac{\sin(kr + \delta_l - \frac{1}{2}\pi l)}{kr} = \\ &= B_l \frac{\exp i(kr + \delta_l - \frac{1}{2}\pi l) - \exp -i(kr + \delta_l - \frac{1}{2}\pi l)}{2ikr}. \end{aligned} \quad (86.2)$$

We recall (see §36) that if the potential energy $U(r)$ is equal to zero over all space, then the set of phase shifts δ_l reduces to zero. The asymptotic expression which we need for ψ for the motion of a particle in the potential field $U(r)$ can be written in the following form:

$$\psi = \sum_{l=0}^{\infty} C_l P_l(\cos \theta) \frac{\exp i(kr + \delta_l - \frac{1}{2}\pi l) - \exp -i(kr + \delta_l - \frac{1}{2}\pi l)}{2ikr}. \quad (86.3)$$

We now have to write expression (83.3) in the form (86.3). This will allow us to relate the coefficients C_l and the phase shifts δ_l to the scattering amplitude $f(\theta)$. Expression (83.3) is most simply brought into the form (86.3) by expanding (83.3) in a series in terms of Legendre polynomials. We need the expansion of the plane wave e^{ikz} only for large distances, which can be found very simply. We write the plane wave in the form

$$e^{ikz} = e^{ikr \cos \theta} = \sum_{l=0}^{\infty} i^l (2l+1) \bar{P}_l(\cos \theta) G_l(r), \quad (86.4)$$

where $G_l(r)$ is an unknown function of the radius. Multiplying this equation by $P_l(\cos \theta) \sin \theta$ and integrating with respect to θ , we find

$$\frac{1}{2i^l} \int_{-1}^{+1} e^{ikrx} P_l(x) dx = G_l(r). \quad (86.5)$$

We have made use of the conditions of orthogonality and normalization of the Legendre polynomials

$$\int_{-1}^{+1} P_l^2(x) dx = \frac{2}{2l+1}.$$

Integrating the left-hand side of (86.5) by parts, we have

$$G_l(r) = \frac{i^{-l}}{2ikr} e^{ikrx} P_l(x) \Big|_{x=-1}^{x=1} + \text{terms of the order of } r^{-2}.$$

Finally, using the known property of the Legendre polynomials $P_l(1) = 1$, $P_l(-1) = (-1)^l$, we obtain for the function $G_l(r)$ at large distances

$$G_l(r) = \frac{\sin(kr - \frac{1}{2}\pi l)}{kr}.$$

Thus the expansion of the plane wave at large distances is written in the form

$$e^{ikz} = \sum_l i^l (2l+1) P_l(\cos \theta) \frac{\sin(kr - \frac{1}{2}\pi l)}{kr}. \quad (86.6)$$

We also expand $f(\theta)$ in a series in terms of the Legendre polynomials

$$f(\theta) = \sum_{l=0}^{\infty} D_l P_l(\cos \theta). \quad (86.7)$$

Substituting series (86.6) and (86.7) into (83.3) and equating the expression found and the asymptotic expression (86.3), we have

$$\begin{aligned}
\sum_l C_l \frac{P_l(\cos \theta)}{2ikr} [\exp i(kr - \frac{1}{2}\pi l + \delta_l) - \exp -i(kr - \frac{1}{2}\pi l + \delta_l)] = \\
= \sum_l \left[\frac{i^l(2l+1)}{2ikr} [\exp i(kr - \frac{1}{2}\pi l) - \exp -i(kr - \frac{1}{2}\pi l)] + D_l \frac{e^{ikr}}{r} \right] P_l(\cos \theta).
\end{aligned} \quad (86.8)$$

For eq. (86.8) to be fulfilled for arbitrary values of the angle θ it is necessary that the coefficients of the polynomials P_l on each side be equal to each other. Equating these coefficients we find

$$\begin{aligned}
\frac{C_l}{2ikr} [\exp i(kr - \frac{1}{2}\pi l + \delta_l) - \exp -i(kr - \frac{1}{2}\pi l + \delta_l)] = \\
= \frac{i^l(2l+1)}{2ikr} [\exp i(kr - \frac{1}{2}\pi l) - \exp -i(kr - \frac{1}{2}\pi l)] + D_l \frac{e^{ikr}}{r}.
\end{aligned} \quad (86.9)$$

This relation must be fulfilled for any arbitrary value of the radius r . This means that the coefficients of exponentials with the same indices must be equal to each other. Hence we find the following relation between the coefficients:

$$\begin{aligned}
C_l &= i^l(2l+1) \exp(i\delta_l), \\
i^l(2l+1) + 2ikD_l \exp(\frac{1}{2}i\pi l) &= C_l \exp(i\delta_l).
\end{aligned} \quad (86.10)$$

Finding D_l from this and substituting it into expansion (86.7), we find for the scattering amplitude the expression

$$f(\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1) [e^{2i\delta_l} - 1] P_l(\cos \theta). \quad (86.11)$$

Consequently, the differential cross section will be equal to

$$d\sigma = \frac{1}{4k^2} \left| \sum_{l=0}^{\infty} (2l+1) (e^{2i\delta_l} - 1) P_l(\cos \theta) \right|^2 d\Omega. \quad (86.12)$$

We find the total cross section by integrating (86.12) and taking into account the orthogonality relations for the Legendre polynomials. A simple calculation gives

$$\sigma = \sum_{l=0}^{\infty} \frac{4\pi}{k^2} (2l+1) \sin^2 \delta_l. \quad (86.13)$$

We see that the differential cross section and the total cross section for the scattering of a particle in a given force field is expressed in terms of the set of phase shifts δ_l . Hence it follows that for the calculation of scattering cross sections it is necessary to find the solution of the Schrödinger equation (35.8) for a particle moving in the given force field. Defining the form of the solution for large distances and comparing it with (86.2), we find δ_l .

The exact solution of the Schrödinger equation makes it possible to find the infinite set of phase shifts δ_l and, consequently, the value of the scattering cross section. The exact or partial wave theory of scattering was first developed by Rayleigh, who studied the scattering of sound waves. Faxen and Holtsmark were the first to use Rayleigh's method for solving the problems of quantum mechanics.

From (86.13) it is seen that the total cross section can be written in the form of a sum of the so-called partial cross sections

$$\sigma = \sum_{l=0} \sigma_l, \quad \sigma_l = \frac{4\pi}{k^2} (2l+1) \sin^2 \delta_l.$$

Each of the partial cross sections corresponds to taking into account one of the terms of the series (86.2)

$$B_l P_l(\cos \theta) \frac{\sin(kr - \frac{1}{2}\pi l + \delta_l)}{kr}.$$

It is clear that it describes a state of the particle with definite angular momentum $L^2 = \hbar^2 l(l+1)$. For this reason a notation analogous to the notation of atomic terms is adopted in scattering theory. For example, to $l=0$ there corresponds s-wave scattering, which is characterized by the partial cross section σ_0 ; to $l=1$ there corresponds p-wave scattering with the partial cross section σ_1 and so on.

The total particle flux through an arbitrary surface surrounding the scattering centre for a particle in the state with angular momentum L , is equal to zero. It could be calculated by the general formula (7.3). However, this can also be seen without carrying out the calculation, on the basis of the general theorem presented in §7. There it was pointed out that the total flux is always equal to zero in the case of a real wave function. In our case this is so, since the wave function is expressed by formula (86.2).

The equality to zero of the total flux of scattered particles has an obvious meaning; it means the law of conservation of the number of particles in the process of scattering. It is important to note that the conservation law holds for particles of each value of l separately. We shall come back to the discussion of this fact in §91.

Finding the values of all the phase shifts δ_l is, as a rule, a very complex problem. Furthermore, the practical value of formulae written in the form of a series is not great if the series does not have a sufficiently rapid convergence. We cannot dwell here on the problems of convergence of the series (86.12) and (86.13) and shall give only the final result*.

For the convergence of series (86.13) it is necessary that the potential energy $U(r)$ should decrease at large distances more rapidly than r^{-n} , where $n > 2$. Further, the series for the differential cross section diverges for $\theta = 0$, if $U(r)$ has the form r^{-n} , where $n \leq 3$ at large distances. When $r \rightarrow 0$, $U(r)$ must increase more slowly than r^{-2} .

The practical value of formulae (86.12) and (86.13) for the scattering cross section becomes greater, the smaller the number of terms of the series which play an essential role. Simple reasoning shows that as the energy of the particle increases the number of phase shifts δ_l which must be taken into account in series (86.12) and (86.13) increases.

Indeed, let R be the radius of the sphere in which the interaction energy is substantially different from zero. For a sufficiently rapid decrease of $U(r)$ the introduction of such a quantity is always possible. The wave function R_l has its first maximum at the distance r defined by the relation $kr \sim l$. At the next maximum R_l has a considerably smaller value because of the decrease of the factor r^{-1} .

For small values of r the wave function is also small. Thus the wave function R_l has its basic value for $r \sim l/k$. If $r \sim l/k > R$, then the wave function is small in the interaction sphere. But in this case the scattering amplitude will also be small. Thus only those particles for which $l/k \leq R$ undergo effective scattering.

The angular momentum l of the effectively scattered particles increases with increasing energy of the particle. For small energies the number of terms which must be taken into account in the series (86.12) and (86.13) is relatively small. Therefore partial wave scattering theory is particularly important for the study of the scattering of slow particles. This qualitative reasoning can be replaced by a quantitative rule, which we shall give without proof.

* L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

If a classical particle having momentum p and impact parameter

$$\rho_l = \frac{\hbar [l(l+1)]^{\frac{1}{2}}}{p} = \frac{[l(l+1)]^{\frac{1}{2}}}{k}, \quad (86.14)$$

in moving does not penetrate the region where the potential energy of interaction differs considerably from zero, then the phase δ_l corresponding to the angular momentum $\hbar^2 l(l+1)$ is small*.

We apply this rule to the investigation of the scattering of a slow particle. Let the scattering centre produce a field effective to the range R . By slow particles we shall mean particles with quantum number k for which $kR \ll 1$.

In this case

$$\rho_l > R. \quad (86.15)$$

For all values $l > 0$ all phase shifts, except δ_0 , are small. We thus see that only s-wave scattering is important for the scattering of slow particles.

The differential cross section is then equal to

$$d\sigma = \frac{1}{4k^2} |e^{2i\delta_0} - 1|^2 d\Omega = \frac{\sin^2 \delta_0}{k^2} d\Omega, \quad (86.16)$$

since $P_0(\theta) = 1$.

The cross section for s-wave scattering does not depend on the scattering angle. This means that the scattering is spherically symmetric. As the energy of the particle increases phase shifts of higher order begin to play a role and the scattering progressively assumes an ever more asymmetric character.

For large energies the cross section becomes substantially different from zero only for very small angles θ . This can best be seen by means of the Born approximation (84.3). For large energies the vector \mathbf{K} is large, the integral rapidly oscillates, and hence the cross section is small. For $\theta = 0$ the vector \mathbf{K} is equal to zero, there is no oscillation, and the cross section is large. Finally, we note that partial wave scattering theory in the form in which it has been described here is inapplicable to scattering in the Coulomb field. The wave function in this case does not have the asymptotic form (83.3). This fact is associated with the very slow decrease of the Coulomb potential as a function of the distance. This case requires particular consideration**.

* The derivation of this statement is given in the book: N.F.Mott and H.S.W.Massey, *The theory of atomic collisions* (Clarendon Press, Oxford, 1965).

** For the exact solution of this problem see, for example, L.D.Landau and E.M. Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

§87. Scattering by a spherical potential well (the concept of resonance scattering)

As an example of the use of partial wave theory we shall consider the scattering of a particle in a potential field which we define in the following way:

$$\begin{aligned} U &= -U_0 & \text{for } r < R, \\ U &= 0 & \text{for } r > R. \end{aligned} \quad (87.1)$$

For simplicity we confine ourselves to the case where the scattered particle has a small energy, i.e. $kR \ll 1$. In this case, as we know, s-wave scattering is important, and we need only determine the phase shift δ_0 . In the case of the potential field given by formula (87.1) the solution of the problem offers no difficulty. By means of the relations already found we can also illustrate a very interesting phenomenon occurring in the scattering process, so-called resonance scattering. It consists in the fact that the scattering cross section under certain conditions turns out to be very large. This effect occurs when there exists an energy level in the potential field close to zero and the energy of the particle to be scattered is sufficiently small. We write the wave function in the form $\psi = A_0 R_0(r) = \chi/r$. The function $\chi(r)$ satisfies the equation

$$\frac{d^2\chi}{dr^2} + k^2\chi = 0 \quad \text{for } r > R, \quad (87.2)$$

$$\frac{d^2\chi}{dr^2} + \beta^2\chi(r) = 0 \quad \text{for } r < R, \quad (87.3)$$

where

$$\beta^2 = \frac{2m(E + U_0)}{\hbar^2}.$$

The form of the function χ for $r > R$ is easily obtained from the solution of eq. (87.2)

$$\chi = C \sin(kr + \delta_0). \quad (87.4)$$

In the general case the function χ has the form (87.4) only for large distances (see formula (86.2)). However, in our case, by virtue of the sharp boundary of the potential energy, the function R_0 has the form (86.2) for all distances $r > R$.

For $r < R$ we obtain

$$\chi = A \sin \beta r + B \cos \beta r,$$

where A and B are constants. The function R_0 must remain finite for $r \rightarrow 0$. Hence the coefficient B must be set equal to zero. Thus we get

$$\chi = A \sin \beta r \quad \text{for} \quad r < R.$$

The function ψ and its first derivative must be continuous at the point $r = R$. These two relations are conveniently replaced by the equality of the logarithmic derivatives. We then find

$$\beta \cotan \beta R = k \cotan (kR + \delta_0). \quad (87.5)$$

We have obtained a transcendental equation for the phase shift δ_0 . We first assume that δ_0 is small. Then $\cotan (kR + \delta_0)$ can be expanded in a series in terms of the small argument $kR + \delta_0$. As a result we have

$$\beta \cotan \beta R = \frac{k}{kR + \delta_0},$$

whence we can find the phase shift δ_0 which is equal to

$$\delta_0 = \frac{k}{\beta \cotan \beta R} - kR. \quad (87.6)$$

We see from relation (87.6) that the phase shift will indeed be considerably smaller than unity if the following relation is fulfilled:

$$\frac{k}{\beta \cotan \beta R} \ll 1. \quad (87.7)$$

The differential cross section can easily be found by making use of formula (86.16) and recalling that $\delta_0 \ll 1$. It has the form

$$d\sigma = \frac{\delta_0^2}{k^2} d\Omega = \frac{1}{k^2} \left(\frac{k}{\beta \cotan \beta R} - kR \right)^2 d\Omega.$$

However, a form of the potential well for which $\beta \cotan \beta R$ approaches zero is possible. In this case inequality (87.7) is violated, and the phase shift δ_0 is large. To find the conditions under which δ_0 is large, we shall establish the relation between the quantity $\beta \cotan \beta R$ involved in formula (87.5) and the energy level of the particle in a bound state. In §37 we have obtained for the energy levels of a particle in a potential well formula (37.6)

$$\frac{2m(U_0 - \epsilon)}{\hbar^2} \cotan^2 \left(\frac{2m(U_0 - \epsilon)R^2}{\hbar^2} \right)^{\frac{1}{2}} = \frac{2m\epsilon}{\hbar^2}. \quad (87.8)$$

ϵ is the energy level of the particle in the well. If the energy level of the particle in the well is close to zero, i.e. if $\epsilon \ll U_0$, then relation (87.8) may be rewritten in the form

$$\frac{2mU_0}{\hbar^2} \cotan^2 \left(\frac{2mU_0 R^2}{\hbar^2} \right)^{\frac{1}{2}} = \frac{2m\epsilon}{\hbar^2}. \quad (87.9)$$

In the case being considered the energy of the scattered particle is also small ($E \ll U_0$), hence relation (87.9) can be written in the form

$$\beta^2 \cotan^2 \beta R = \frac{2m\epsilon}{\hbar^2}. \quad (87.10)$$

Thus we see that the increase of δ_0 is associated with the presence of an energy level ϵ close to zero.

We now turn to the calculation of the phase shift in the case where relation (87.7) is violated and δ_0 is large. We denote this value of the phase shift by δ_{0r} . We find δ_{0r} again from relations (87.5). For this we expand $\cotan(kR + \delta_{0r})$ in a series in terms of the small parameter kR and restrict ourselves to the zero order term of the expansion. Then we obtain

$$\beta \cotan \beta R = k \cotan \delta_{0r}.$$

Squaring this relation and making use of formula (87.10), we find

$$\cotan^2 \delta_{0r} = \frac{\epsilon}{E}. \quad (87.11)$$

We now see that the phase shift δ_{0r} will not be a small quantity if $\epsilon < E$.

We find the scattering cross section by means of the general formula (86.16). In this case we have

$$\sigma_r = \frac{4\pi \sin^2 \delta_{0r}}{k^2} = \frac{2\pi \hbar^2}{m(E + \epsilon)}. \quad (87.12)$$

This expression is called the Wigner formula. It is easily seen that the cross section in the case of a resonance is considerably larger than in the absence of a resonance. The ratio of the cross section is equal to

$$\frac{\sigma_r}{\sigma} = \frac{\sin^2 \delta_{0r}}{\delta_0^2}.$$

Since $\delta_0 \ll 1$, and $\sin \delta_{0r}$, as is seen from formula (87.11), for $\epsilon \approx E$ is close to one, then it is evident that

$$\frac{\sigma_r}{\sigma} \gg 1.$$

We have obtained formula (87.12) for a particular form of the potential energy. It should be stressed, however, that the dependence of the cross section on ϵ (87.12) is general and is not related to the actual form of the potential energy*.

Resonance scattering also occurs in the case where the system does not have a real level close to zero but the configuration of the field is similar to that for which such a level appears. In such a situation the function $\cotan \beta R$ is positive, whereas for the real level we necessarily have $\cotan \beta R < 0$ (see §37). Relation (87.10) contains $\cotan^2 \beta R$ and hence is fulfilled independently of the sign of the function $\cotan \beta R$. In the case where $\cotan \beta R > 0$, scattering takes place at the virtual level, not at the real level.

By means of the relations obtained above one can also easily find the differential cross section for scattering by a potential barrier, i.e. by a potential field having the following form:

$$U = 0 \quad \text{for } r > R,$$

$$U = |U_0| \quad \text{for } r < R.$$

For this it is sufficient to carry out the replacement $\beta \rightarrow i\beta$. Then for the differential cross section we obtain

$$d\sigma = \frac{1}{|\beta|^2} (\tanh |\beta|R - |\beta|R)^2 d\Omega. \quad (87.13)$$

Formula (87.13) is simplified in the case of an infinitely high potential barrier $U_0 \rightarrow \infty$. In this case we find the following expression for the total cross section:

$$\sigma = 4\pi R^2. \quad (87.14)$$

It is interesting to note that the scattering cross section is larger in this case than the geometric size of the scatterer, by a factor of four.

* For a more general derivation of the formula for resonance scattering see L.D. Landau and E.M. Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

§88. The elastic scattering of identical particles

Up till now we have assumed that the scattered particle and the target are different particles. We now consider the case where the scattered and target particles are identical. As we shall now see, the identity of the particles has an essential effect on the scattering process. We shall begin with the consideration of particles of zero spin. We first suppose that the identical particles move towards each other with equal velocities. In this case the centre of mass of the system is at rest and the wave function will, in correspondence with (14.14), have the form

$$\psi = \psi(x, y, z)$$

and will depend only on the relative coordinates. The wave function $\psi_0(x, y, z)$ satisfies eq. (14.11)

$$\left[-\frac{\hbar^2}{2\mu} \nabla^2 + U(r) \right] \psi = E\psi(x, y, z). \quad (88.1)$$

The reduced mass of two identical particles is equal to $\mu = \frac{1}{2}m$. We cannot for our case write the wave function in the form

$$\psi = e^{ikz} + \frac{f(\theta)}{r} e^{ikr},$$

since this function does not satisfy the symmetry requirements.

As a matter of fact according to (14.6) there corresponds to the exchange of two particles (i.e. to the replacement $x_1 \rightarrow x_2, y_1 \rightarrow y_2, z_1 \rightarrow z_2$) the transformation $\mathbf{r} \rightarrow -\mathbf{r}$. Here the modulus of the vector \mathbf{r} does not change, and the angle θ is replaced by $\pi - \theta$. Taking into account this last transformation, it is easily found that the symmetrized wave function must have the form

$$\psi_s = e^{ikz} + e^{-ikz} + \frac{e^{ikr}}{r} [f(\theta) + f(\pi - \theta)]. \quad (88.2)$$

A diverging wave again describes the scattered particles. The differential scattering cross section is now given by the expression

$$d\sigma = |f(\theta) + f(\pi - \theta)|^2 d\Omega = |f(\theta) + f(\pi - \theta)|^2 \sin \theta d\theta d\varphi. \quad (88.3)$$

Thus we have found the differential cross section for the process in which one of the colliding identical particles is scattered at an angle θ with respect to the direction of its initial flight.

From formula (88.3) it follows that the number of particles scattered at angle θ and at angle $\pi - \theta$ is the same. If one of the particles was at rest

before the collision, then the differential cross section in this system of coordinates can be found in the following way. In the system of coordinates in which the centre of mass is at rest the differential cross section is given by expression (88.3). The transition to the laboratory system is carried out by means of formulae (83.2). In the case given the mass of the particles is the same, and we obtain

$$\tan \vartheta_1 = \frac{\sin \theta}{1 + \cos \theta} = \tan \frac{1}{2} \theta$$

and, correspondingly, $\vartheta_1 = \frac{1}{2} \theta$.

Expressing the differential cross section as a function of the angle ϑ_1 , we find

$$\begin{aligned} d\sigma &= |f(2\vartheta_1) + f(\pi - 2\vartheta_1)|^2 4 \cos \vartheta_1 \sin \vartheta_1 d\vartheta_1 d\varphi_1 = \\ &= |f(2\vartheta_1) + f(\pi - 2\vartheta_1)|^2 4 \cos \vartheta_1 d\Omega_1, \end{aligned} \quad (88.4)$$

where $d\Omega_1$ is a solid angle element in the laboratory system.

Expression (88.4) gives the differential cross section for the process in which one of the particles is scattered into the solid angle element $d\Omega_1$. Since the two particles are identical, the question as to which one of the particles entered $d\Omega_1$, that which was initially moving or that which was initially at rest, makes no physical sense.

As an example of the application of formula (88.4) let us consider the collision of two identical particles in which the interacting energy has the simple form

$$\begin{aligned} U &= U_0 \quad \text{for } r < R, \\ U &= 0 \quad \text{for } r > R. \end{aligned}$$

We suppose that before the collision one of the particles was at rest while the other was moving sufficiently slowly that the relation $kR \ll 1$ was fulfilled. In this case $\delta_l \ll 1$ and in accord with (86.11) the scattering amplitude can be written in the form $f(\theta) = \delta_0/k$.

For the differential scattering cross section in the laboratory system we obtain

$$d\sigma = |f(2\vartheta_1) + f(\pi - 2\vartheta_1)|^2 4 \cos \vartheta_1 d\Omega_1 = \frac{16\delta_0^2}{k^2} \cos \vartheta_1 d\Omega_1.$$

Thus we see that if in the centre-of-mass system the scattering is spherically symmetric, then in the laboratory system the differential cross section is proportional to the cosine of the scattering angle.

The theory of the scattering of identical particles with spin different from zero is constructed according to the same scheme as for spinless particles. For concreteness we assume that both colliding particles have spin $\frac{1}{2}$. The generalization of the theory to the case of arbitrary spin offers no difficulty.

We consider the collision of two identical particles in the centre-of-mass system in the case where the total spin of the system is equal to zero (i.e. the spins of the particles are antiparallel). Then the spin part of the wave function must be antisymmetric and, consequently, the coordinate part must be symmetric. In other words, the coordinate part of the wave function can, as in the case of spinless particles, be written in the form

$$\psi_s = e^{ikz} + e^{-ikz} + \frac{e^{ikr}}{r} [f(\theta) + f(\pi - \theta)] . \quad (88.5)$$

Correspondingly we have for the differential scattering cross section

$$d\sigma_s = |f(\theta) + f(\pi - \theta)|^2 d\Omega . \quad (88.6)$$

If the total spin is equal to one (i.e. the spins are parallel), then the spin part of the wave function is symmetric, and the coordinate part is antisymmetric. Hence for this case we can write the following asymptotic expression:

$$\psi_a = e^{ikz} - e^{-ikz} + \frac{e^{ikr}}{r} [f(\theta) - f(\pi - \theta)] . \quad (88.7)$$

Then for the differential cross section we obtain

$$d\sigma_a = |f(\theta) - f(\pi - \theta)|^2 d\Omega . \quad (88.8)$$

We have considered above processes in which the scattered particles had a definite spin orientation. However, in scattering, particles are often in a state with indefinite spin. In this case one is usually interested in the mean cross section which is obtained by averaging over all possible spin states. The mean cross section for particles with spin $\frac{1}{2}$ can easily be found as follows. The colliding particles can be in four states: in one state with spin 0 and in three states with spin 1 (three possible projections on the z-axis). Since all these states are equally probable, the state with spin 0 has a statistical weight equal to $\frac{1}{4}$, and the weight of the state with spin 1 is equal to $\frac{3}{4}$. Hence the mean differential cross section can be written in the form

$$d\sigma = \frac{1}{4} d\sigma_s + \frac{3}{4} d\sigma_a . \quad (88.9)$$

As an example let us consider the scattering of two slow identical particles with spin $\frac{1}{2}$, for which the interaction energy can be written in the form

$$U = U_0 \quad \text{for } r < R,$$

$$U = 0 \quad \text{for } r > R.$$

In the case of parallel spins the scattering cross section given by expression (88.8) turns out to be equal to zero

$$d\sigma_a = |f(\theta) - f(\pi - \theta)|^2 d\Omega = 0.$$

Consequently, the scattering of particles with parallel spins is associated with effects of higher orders, i.e. with p-wave, d-wave etc. scattering. The cross section for the scattering of particles with antiparallel spins at small energies is the same as for particles with spin zero,

$$d\sigma_s = \frac{4\delta_0^2}{k^2} d\Omega.$$

The mean cross section according to formula (88.9) is given by the expression

$$d\sigma = \frac{\delta_0^2}{k^2} d\Omega.$$

Thus we see that taking into account the identity of the particles leads to the appearance of a basic dependence of the scattering cross section on the mutual orientation of their spins.

The transition from cross sections calculated in the centre-of-mass system to cross sections calculated in the laboratory system is carried out in the same way as for spinless particles.

§89. The effect of polarization in scattering processes

All the results obtained up to now apply to the scattering of beams in which all the particles are in one and the same state, i.e. are described by one and the same wave function. However, the particles of a beam can be in different spin states. We shall now confine ourselves to considering beams made up of particles with spin $\frac{1}{2}$ scattered by non-polarized targets. As is known, each of the particles of the beam is described by a two-component spinor.

In §61 it was shown that an arbitrary state of a particle is at the same time a state with a definite projection of the spin on a certain direction in space. In other words, for a state with an indefinite z-component of the spin one can always find some z' -axis with respect to which the given state will be a state

with a definite spin projection. Consequently, we see that if the beam consists of particles which are in the same state, then it will be fully polarized along a certain direction. If the beam is partially polarized, then the particles are described by different spinors. In this case the beam cannot be described by means of a wave function, and we have a mixture of states (see §23). Nevertheless, for the description of the spin properties of the particles of the beam one can introduce a function φ defined by the formula*

$$\varphi = c_1 \varphi_1 \epsilon_1 + c_2 \varphi_2 \epsilon_2 + \dots$$

The summation is carried out over the spin states of the particles of the beam. We denote the spinor which describes the group of particles in the k th spin state by φ_k . The coefficient c_k determines the weight of this state. It is proportional to the number of particles in a given group. The quantities ϵ_i and ϵ_k , satisfying the condition $\epsilon_i^2 = 1$ and $\epsilon_i \epsilon_k = 0$, are introduced in order to eliminate the interference between wave functions of particles which are in different spin states in the quadratic expressions defining the mean values. We define the polarization vector as the spin vector averaged over the beam:

$$\mathbf{P} = \bar{\boldsymbol{\sigma}} = \frac{\varphi^\dagger \boldsymbol{\sigma} \varphi}{\varphi^\dagger \varphi} = \frac{\sum_n |c_n|^2 \varphi_n^\dagger \boldsymbol{\sigma} \varphi_n}{\sum_n |c_n|^2} = \frac{\sum \varphi_n'^\dagger \boldsymbol{\sigma} \varphi_n'}{\sum \varphi_n'^\dagger \varphi_n'}, \quad (89.1)$$

where $\varphi_n' = c_n \varphi_n$.

Formula (89.1) has a simple meaning: $(\varphi_n^\dagger \boldsymbol{\sigma} \varphi_n)$ represents the mean value of the spin vector in the n th state, and the ratio $|c_n|^2 / \sum_n |c_n|^2$ determines the probability of realization of the n th state in the beam. This probability is equal to N_n/N , where N_n is the number of particles in the n th state, and N is the total number of particles in the beam.

We write $c_n \varphi_n$ in the form

$$c_n \varphi_n = \begin{pmatrix} u_n \\ v_n \end{pmatrix}. \quad (89.2)$$

Substituting expression (89.2) into (89.1), we easily find the components of the polarization vector

* L. Wolfenstein, Phys. Rev. 75 (1943) 1664.

$$\begin{aligned}
 P_x &= \frac{2 \operatorname{Re} \sum_n u_n^* v_n}{\sum_n (|u_n|^2 + |v_n|^2)}, \\
 P_y &= \frac{2 \operatorname{Im} \sum_n u_n^* v_n}{\sum_n (|u_n|^2 + |v_n|^2)}, \\
 P_z &= \frac{\sum_n (|u_n|^2 - |v_n|^2)}{\sum_n (|u_n|^2 + |v_n|^2)}.
 \end{aligned} \tag{89.3}$$

If one half of the particles constituting the beam is polarized in some direction, for example in the positive direction of the z -axis, and the other half is polarized in the opposite direction, then the polarization vector \mathbf{P} of the beam will be equal to zero. Indeed, one group of particles is described by the spin functions.

$$c_1 \varphi_1 = \begin{pmatrix} c_1 \\ 0 \end{pmatrix},$$

while the other group has the spin functions

$$c_2 \varphi_2 = \begin{pmatrix} 0 \\ c_2 \end{pmatrix}, \quad |c_1|^2 = |c_2|^2.$$

Substituting these values into (89.3), we find that the polarization vector is equal to zero.

We now turn to the case of the scattering of particles with spin $\frac{1}{2}$ by a target with spin 0. Then the wave function ψ , describing the process of elastic scattering, at large distances has the form

$$\psi = \varphi e^{ikz} + \frac{e^{ikr}}{r} f \varphi. \tag{89.4}$$

Here φ is the spinor characterizing the state of the incident particle, and f is a certain two-row matrix depending on the scattering angles. Let us establish

the general form of this matrix. First of all we note that any two-row matrix can be expressed in terms of a unit matrix and the Pauli matrices $\sigma_x, \sigma_y, \sigma_z$, since the matrices mentioned make up a complete system (see §60). Correspondingly we have

$$f = g(\theta)I + \mathbf{h}(\theta) \cdot \boldsymbol{\sigma}. \quad (89.5)$$

The further form of the functions g and \mathbf{h} can be obtained from the following considerations. The laws of transformation of the first and second terms of formula (89.4) must be the same under spatial rotations and reflections. Since the first term transforms as a spinor, the second term in this formula must also have the character of a spinor. Hence it follows that the function g must be a scalar. Since the operator $\boldsymbol{\sigma}$ transforms as a pseudovector, \mathbf{h} also must be a pseudovector. On the other hand, the pseudovector \mathbf{h} depends on the quantities which characterize the scattering process, and can be defined by only two vectors \mathbf{k}_0 and \mathbf{k}_1 (the wave vectors of the particle before and after scattering). From these two vectors one can construct the single unit pseudovector

$$\mathbf{n} = \frac{\mathbf{k}_0 \times \mathbf{k}_1}{|\mathbf{k}_0 \times \mathbf{k}_1|}.$$

Hence $\mathbf{h} = h(\theta)\mathbf{n}$ where $h(\theta)$ is a scalar.

Finally we obtain

$$f = g(\theta)I + \mathbf{n} \cdot \boldsymbol{\sigma} h(\theta). \quad (89.6)$$

Correspondingly, the elastic scattering cross section has the form

$$\frac{d\sigma}{d\Omega} = \varphi^\dagger f^\dagger f \varphi = |g|^2 + |h|^2 + 2 \operatorname{Re}(g^* h) \boldsymbol{\mu} \cdot \mathbf{n}. \quad (89.7)$$

where $\boldsymbol{\mu} = \varphi^\dagger \boldsymbol{\sigma} \varphi$.

We average expression (89.7) over the spin states of the particles of the incident beam. Then, making use of (89.1), we find

$$\begin{aligned} \frac{d\sigma}{d\Omega} &= |g|^2 + |h|^2 + 2 \operatorname{Re}(g^* h) \mathbf{P}_{\text{inc}} \cdot \mathbf{n} = \\ &= (|g|^2 + |h|^2) \left(1 + \frac{2 \operatorname{Re}(g^* h) \mathbf{P}_{\text{inc}} \cdot \mathbf{n}}{|g|^2 + |h|^2} \right). \end{aligned} \quad (89.8)$$

where \mathbf{P}_{inc} is the polarization vector of the incident beam.

If the incident beam is not polarized ($\mathbf{P}_{\text{inc}} = 0$), then the differential cross section is equal to

$$\frac{d\sigma}{d\Omega} = |g|^2 + |h|^2 \quad (89.9)$$

We now turn to the investigation of the state of the scattered beam. We stress that a polarization of the beam can arise after scattering even in the case where the incident beam was not polarized. From general considerations it is easy to indicate the direction of polarization of the scattered beam. As a matter of fact, the polarization is described by the pseudovector \mathbf{P} , which can be oriented only in the direction of the pseudovector \mathbf{n} . Consequently, for the scattered beam, which was not polarized before scattering, we have

$$\mathbf{P}_{\text{scat}} = P_{\text{scat}} \mathbf{n}. \quad (89.10)$$

We shall determine the value of the polarization of the scattered beam. Based on definition (89.1), we have

$$\mathbf{P}_{\text{scat}} = \frac{\sum_n (f\varphi_n)^\dagger \boldsymbol{\sigma} (f\varphi_n)}{\sum_n (f\varphi_n)^\dagger (f\varphi_n)} = \frac{\sum_n \varphi_n^\dagger f^\dagger \boldsymbol{\sigma} f \varphi_n}{\sum_n \varphi_n^\dagger f^\dagger f \varphi_n}. \quad (89.11)$$

Since by assumption the incident beam is not polarized, it can be represented in the form of two beams consisting of the same number of particles but with oppositely directed spins. Then the summation over n reduces to the summation over two states characterized by oppositely directed spins.

Consequently, we have

$$\mathbf{P}_{\text{scat}} = \frac{\sum_{i=1}^2 \varphi_i^\dagger f^\dagger \boldsymbol{\sigma} f \varphi_i}{\sum_{i=1}^2 \varphi_i^\dagger f^\dagger f \varphi_i}.$$

We see that to calculate the polarization it is necessary to find the sums of the diagonal elements (traces) of the matrices of certain operators. In the notation of § 45 this last formula can be rewritten in the form

$$\mathbf{P}_{\text{scat}} = \frac{\text{Tr} f^\dagger \boldsymbol{\sigma} f}{\text{Tr} f^\dagger f}. \quad (89.12)$$

We calculate first $\text{Tr} f^\dagger \boldsymbol{\sigma} f$. Making use of expression (89.6), we have

$$\text{Tr} f^\dagger \boldsymbol{\sigma} f = \text{Tr} \{ [g^* I + h^* (\mathbf{n} \cdot \boldsymbol{\sigma})] \boldsymbol{\sigma} [g I + h (\mathbf{n} \cdot \boldsymbol{\sigma})] \}.$$

It is easily seen from formulae (60.15) and (60.16) that $\text{Tr } \sigma_i = 0$ ($i=1,2,3$). From relations of the type $\sigma_x \sigma_y = i\sigma_z$ it follows that

$$\text{Tr } \sigma_i \sigma_k = 0 \quad (i \neq k).$$

Since $\sigma_i^2 = I$, then $\text{Tr } \sigma_i^2 = 2$ ($i=1,2,3$).

Using these relations, we obtain

$$\begin{aligned} \text{Tr } f^\dagger \sigma f &= \text{Tr} [g^* h \sigma(\mathbf{n} \cdot \sigma) + h^* g (\mathbf{n} \cdot \sigma) \sigma] = \\ &= \text{Tr} (g^* h + h^* g) (\sigma_x^2 n_x i + \sigma_y^2 n_y j + \sigma_z^2 n_z k) = 4n \text{Re} (g^* h). \end{aligned}$$

By means of analogous calculations we find

$$\text{Tr } f^\dagger f = 2(|h|^2 + |g|^2).$$

Thus the polarization vector of the scattered beam has the form

$$\mathbf{P}_{\text{scat}} = \frac{2 \text{Re} (g^* h)}{|g|^2 + |h|^2} \mathbf{n}; \quad \mathbf{n} = \frac{\mathbf{k} \times \mathbf{k}_1}{|\mathbf{k} \times \mathbf{k}_1|}. \quad (89.13)$$

Making use of (89.13) and (89.8), we express the scattering cross section in terms of the polarization vector \mathbf{P}_{scat} :

$$\frac{d\sigma}{d\Omega} = (|g|^2 + |h|^2)(1 + \mathbf{P}_{\text{inc}} \cdot \mathbf{P}_{\text{scat}}), \quad (89.14)$$

where \mathbf{P}_{scat} is the polarization vector of the beam of scattered particles in the case where the beam was not polarized before scattering.

Thus we see that the scattering cross section depends on the polarization of the incident and scattered beams. Experimentally such dependences can be observed in experiments on double scattering. An unpolarized beam of particles (fig. V.26) becomes polarized after scattering. Then the polarized beam of particles falls on the second scatterer. In this case the cross section for scattering to the left (vector \mathbf{k}_2) turns out to be different from the cross section for scattering to the right (vector \mathbf{k}'_2).

For simplicity we assume that all the vectors \mathbf{k} , \mathbf{k}_1 , \mathbf{k}_2 and \mathbf{k}'_2 lie in one plane. The vector \mathbf{n} , characterizing the polarization after the first scattering, is directed upwards perpendicular to the plane of the drawing. The vectors \mathbf{P}_{scat} involved in formula (89.14) have opposite directions for beams scattered a second time to the left and to the right because of the different directions of the vectors \mathbf{k}_2 and \mathbf{k}'_2 . Thus the cross section for the beam scattered to the left is equal to

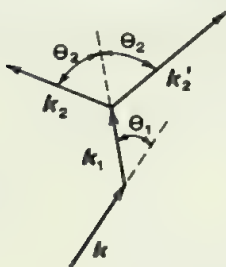


Fig. V.26

$$\frac{d\sigma}{d\Omega} = (|g|^2 + |h|^2) [1 + P_{\text{scat}}(\theta_1) P_{\text{scat}}(\theta_2)] . \quad (89.15)$$

Correspondingly, for the beam scattered to the right we have

$$\frac{d\sigma}{d\Omega} = (|g|^2 + |h|^2) [1 - P_{\text{scat}}(\theta_1) P_{\text{scat}}(\theta_2)] . \quad (89.16)$$

We see that the ratio of the number of particles scattered to the left and to the right is determined by the polarization P_{scat} . We have

$$R = \frac{1 + P_{\text{scat}}(\theta_1) P_{\text{scat}}(\theta_2)}{1 - P_{\text{scat}}(\theta_1) P_{\text{scat}}(\theta_2)} . \quad (89.17)$$

As an example let us consider the scattering of a neutron by a nucleus taking into account the spin-orbit interaction between them. The concept of this interaction was first introduced by Fermi to account for the phenomenon of polarization of fast neutrons. It has the form

$$\hat{H}' = V(r) + W(r) \boldsymbol{\sigma} \cdot \hat{\mathbf{l}} . \quad (89.18)$$

Here $V(r)$ and $W(r)$ are functions depending only on the radius, and \mathbf{l} is the neutron orbital angular momentum operator.

From experiment it follows that parity is conserved in nuclear interactions. Operator (89.18) is constructed in such a way that it automatically satisfies this conservation law. For what follows it is convenient to write the function $W(r)$ in the form

$$W(r) = \frac{1}{r} \frac{d}{dr} Y(r) .$$

Let us find the functions g and h by making use of the Born approximation. As was shown in §84, the amplitude f in this approximation is equal to

$$\begin{aligned} f &= -\frac{m}{2\pi\hbar^2} \int e^{-ik_1 \cdot \mathbf{r}} \hat{H}'(\mathbf{r}) e^{ik_0 \cdot \mathbf{r}} dV = \\ &= -\frac{m}{2\pi\hbar^2} \left[V_{\mathbf{k}_0 - \mathbf{k}_1} + \frac{\hbar}{i} \boldsymbol{\sigma} \cdot \int e^{-ik_1 \cdot \mathbf{r}} \frac{dY}{r dr} [\mathbf{r} \times \nabla] e^{ik_0 \cdot \mathbf{r}} dV \right], \end{aligned}$$

where $V_{\mathbf{k} - \mathbf{k}'}$ is the Fourier component of the function V . By means of elementary transformations we find

$$f = -\frac{m}{2\pi\hbar^2} \left\{ V_{\mathbf{k}_0 - \mathbf{k}_1} - \hbar \boldsymbol{\sigma} \cdot [\mathbf{k}_0 \times \int e^{i(\mathbf{k}_0 - \mathbf{k}_1) \cdot \mathbf{r}} \nabla Y dV] \right\}.$$

Integrating by parts we obtain

$$\begin{aligned} f &= -\frac{m}{2\pi\hbar^2} \left\{ V_{\mathbf{k}_0 - \mathbf{k}_1} + \hbar \boldsymbol{\sigma} \cdot [\mathbf{k}_0 \times \int Y \nabla e^{i(\mathbf{k}_0 - \mathbf{k}_1) \cdot \mathbf{r}} dV] \right\} = \\ &= -\frac{m}{2\pi\hbar^2} \left\{ V_{\mathbf{k}_0 - \mathbf{k}_1} - i\hbar \boldsymbol{\sigma} \cdot [\mathbf{k}_0 \times \mathbf{k}_1] Y_{\mathbf{k}_0 - \mathbf{k}_1} \right\}. \end{aligned} \quad (89.19)$$

Comparing (89.19) and (89.6) we find the functions h and g to be

$$\begin{aligned} g &= -\frac{m}{2\pi\hbar^2} V_{\mathbf{k}_0 - \mathbf{k}_1}, \\ h &= \frac{imk^2 \sin \theta}{2\pi\hbar} Y_{\mathbf{k}_0 - \mathbf{k}_1}. \end{aligned} \quad (89.20)$$

We note that in the first approximation of the perturbation theory considered there is no polarization of the scattered particles. Indeed, substituting relation (89.20) into formula (89.13) we obtain $\mathbf{P}_{\text{scat}} = 0$. However, in a more accurate calculation $\mathbf{P}_{\text{scat}} \neq 0$.

A more general formalism, suitable for the treatment of the scattering of particles by polarized targets, may be found, for example, in the book by Davydov*.

§90. The transition to the classical limit in the quantum scattering formulae

We first of all transform the exact formula for the scattering amplitude into a form convenient for transition to the classical limit.

* A.S. Davydov, *Theorie des Atomkerns* (Deutscher Verlag der Wissenschaften, Berlin, 1963).

If we make use of the expansion of the δ -function in terms of the Legendre polynomials (III.11), then the scattering amplitude (86.11) can be written in the form

$$f(\theta) = \frac{1}{2ik} \sum_{l=0} (2l+1) P_l(\cos \theta) e^{2i\delta_l} - \frac{1}{ik} \delta(1 - \cos \theta). \quad (90.1)$$

For all angles $\theta \neq 0$ formula (90.1) assumes the form

$$f(\theta) = \frac{1}{2ik} \sum_l (2l+1) P_l(\cos \theta) e^{2i\delta_l}. \quad (90.2)$$

In the quasi-classical approximation the radial part of the wave function has the form (43.2)

$$R_l = \frac{A_l}{r p_r^{\frac{1}{2}}} \sin \left[\frac{1}{\hbar} \int_a^r \left(2m[E - U(r)] - \frac{\hbar^2(l + \frac{1}{2})^2}{r^2} \right)^{\frac{1}{2}} dr + \frac{1}{4}\pi \right].$$

The expression for R_l must be understood as an asymptotic expression, i.e. it must be assumed that $r \rightarrow \infty$; a denotes the coordinate of the turning point, where the total energy E is equal to the sum of the potential and centrifugal energies, i.e.

$$E = U(a) + \frac{\hbar^2(l + \frac{1}{2})^2}{2ma^2}.$$

In §43 the condition for definition of the turning point did not involve the centrifugal energy, since the motion was assumed to be one-dimensional.

Comparing the expression for R_l with formula (86.2) we see that the scattering phase shift can be written in the form

$$\delta_l = \frac{1}{\hbar} \int_a^r \left(2m[E - U(r)] - \frac{\hbar^2(l + \frac{1}{2})^2}{r^2} - k^2 \right)^{\frac{1}{2}} dr + \frac{1}{2}\pi(l + \frac{1}{2}) - ka. \quad (90.3)$$

In (90.3) it is necessary to assume that $r \rightarrow \infty$, $l \gg 1$. Then the values of the phase shifts δ_l are very large in absolute magnitude. The formula for the scattering amplitude (90.2) can be simplified by taking into account that in the quasi-classical approximation l must be assumed $\gg 1$. Then for the Legendre polynomials $P_l(\cos \theta)$ one can write an asymptotic expression for $l \gg 1$. They have the form*

* N.N. Lebedev, *Special functions and their application* (Prentice Hall, Englewood Cliffs, N.J., 1965).

$$P_l(\cos \theta) = \frac{1}{i(2\pi l \sin \theta)^{\frac{1}{2}}} \left[e^{i(l+\frac{1}{2})\theta + \frac{1}{4}\pi} - e^{-i(l+\frac{1}{2})\theta - \frac{1}{4}\pi} \right].$$

Then for the scattering amplitude we obtain

$$f \equiv \frac{1}{ik} \sum_{l \geq 1} l P_l(\cos \theta) e^{2i\delta_l} = \frac{1}{k} \sum B(l) (e^{i\alpha(l)} - e^{i\beta(l)}), \quad (90.4)$$

where

$$B(l) = -\left(\frac{l}{2\pi \sin \theta}\right)^{\frac{1}{2}},$$

$$\alpha(l) = 2\delta_l + (l+\frac{1}{2})\theta + \frac{1}{4}\pi,$$

$$\beta(l) = 2\delta_l - (l+\frac{1}{2})\theta - \frac{1}{4}\pi.$$

To obtain $f(\theta)$ it is necessary to sum the series

$$\sum_l B(l) e^{i\alpha(l)} \quad \text{and} \quad \sum_l B(l) e^{i\beta(l)}.$$

We shall consider one of these series, since, as will be clear from what follows, only one of the series has a sum different from zero in a given force field (repulsive or attractive). The quantities $\alpha(l)$, as can be seen from their definition, are large for large l . Hence the terms of the series $\sum B(l) e^{i\alpha(l)}$, containing rapidly oscillating factors, are mutually cancelled. An exception is possible in the case where for a certain value $l = l_0$ the quantity $\alpha(l_0)$ has an extremum, i.e.

$$\left(\frac{d\alpha(l)}{dl}\right)_{l=l_0} = 0. \quad (90.5)$$

Near the extremum the function $\alpha(l)$ changes slowly and the sum of the series reduces to a sum of terms with values of l close to l_0 .

In this case, to carry out the summation the sum can be replaced by an integral. In the integral the integrand is substantially different from zero only for $l \approx l_0$, and the integral can be calculated by the method of steepest descent (see Part III, §20). Thus one can write

$$\sum_l B(l) e^{i\alpha(l)} = B(l_0) e^{i\alpha(l_0)} \int_{-\infty}^{+\infty} e^{i\gamma(l-l_0)^2} dl = e^{i\alpha(l_0)} B(l_0) \int_{-\infty}^{+\infty} e^{-c(l-l_0)^2} dl, \quad (90.6)$$

where

$$\gamma = \frac{1}{2} \left(\frac{d^2 \alpha}{dl^2} \right)_{l=l_0}, \quad c = -i\gamma.$$

The calculation of the integral in (90.6) is carried out directly, and we obtain

$$\sum_l B(l) e^{i\alpha l} = B(l_0) e^{i\alpha(l_0)} \left(\frac{i\pi}{\gamma} \right)^{\frac{1}{2}}. \quad (90.7)$$

By means of this relation the scattering amplitude can be written in the form

$$f(\theta) = \frac{B(l_0)}{k} e^{i\alpha(l_0)} \left(\frac{i\pi}{\gamma} \right)^{\frac{1}{2}}. \quad (90.8)$$

We shall later find the quantity $\alpha(l_0)$, but for the present we shall consider the physical meaning of eq. (90.5). For this we define the derivative $(d\alpha/dl)_{l=l_0}$. By means of relations (90.4) we have

$$\left(\frac{d\alpha}{dl} \right)_{l=l_0} = 2 \left(\frac{d\delta_l}{dl} \right)_{l=l_0} + \theta = 0. \quad (90.9)$$

In differentiating it should be recalled that the angle θ is given, and that we determine the cross section for a definite value of the angle. If we differentiate with respect to l and make use of formula (90.3), we then obtain

$$\begin{aligned} \left(\frac{d\delta_l}{dl} \right)_{l=l_0} &= - \int_a^\infty \frac{\hbar(l_0 + \frac{1}{2}) dr}{r^2 [2m(E-U) - \hbar^2(l_0 + \frac{1}{2})^2/r^2]^{\frac{1}{2}}} - \\ &- \left[2m(E-U) - \hbar^2(l_0 + \frac{1}{2})^2/r^2 \right]^{\frac{1}{2}}_{r=a} \left(\frac{da}{dl} \right)_{l=l_0} + k \frac{da}{dl} - k \frac{da}{dl} + \frac{\pi}{2} = \\ &= - \int_a^\infty \frac{\hbar(l_0 + \frac{1}{2}) dr}{r^2 [2m(E-U) - \hbar^2(l_0 + \frac{1}{2})^2/r^2]^{\frac{1}{2}}} + \frac{1}{2}\pi, \end{aligned}$$

since at the turning point $r=a$ the square root reduces to zero. Condition (90.9) assumes the form

$$- \int_a^\infty \frac{\hbar(l_0 + \frac{1}{2}) dr}{r^2 [2m(E-U) - \hbar^2(l_0 + \frac{1}{2})^2/r^2]^{\frac{1}{2}}} + \frac{1}{2}\pi \pm \frac{1}{2}\theta = 0. \quad (90.10)$$

If we carried out the corresponding calculations for the second sum, then

lower sign in (90.10) would correspond to the extremum $\beta(l)$. For brevity the two conditions are combined. Formula (90.10) defines the value of l_0 .

The quantity $\hbar(l_0 + \frac{1}{2}) = L$ represents the angular momentum. After introducing the angular momentum L formula (90.10) can be transformed into the form

$$\int_a^\infty \frac{L dr}{r^2 [2m(E-U) - L^2/r^2]^{\frac{1}{2}}} = \frac{1}{2}(\pi \pm \theta). \quad (90.11)$$

In classical mechanics the angular momentum can be connected with the impact parameter ρ by means of the following relation:

$$L = m\rho v,$$

where v is the velocity of the particle at infinity. Substituting this value for the angular momentum into formula (90.11), we obtain an expression which is exactly the same as the classical relation connecting the impact parameter with the scattering angle θ^*

$$\int_{r=a}^\infty \frac{m\rho v dr}{r^2 [2m(E-U) - (m\rho v/r)^2]^{\frac{1}{2}}} = \frac{1}{2}(\pi \pm \theta). \quad (90.12)$$

The values of the impact parameter ρ is determined by the positive root of eq. (90.12). It is known from mechanics that in a repulsive force field the positive root of this equation exists only for a negative θ . On the contrary, in an attractive force field this root exists for a positive θ .

Let us consider the case of repulsive forces. Then condition (90.9) can be fulfilled only for $\alpha(l)$ but not for $\beta(l)$. Correspondingly only the first of the series in (90.4) has a sum different from zero.

We now go on to the calculation of the cross section. According to formulae (83.5) and (90.8), it is defined by the expression

$$d\sigma = |f(\theta)|^2 d\Omega = \frac{1}{k^2} |B(l_0)|^2 \frac{\pi}{|\gamma|} d\Omega.$$

The quantity γ is defined by expression (90.6). By means of (90.4) and (90.3) we obtain

* See, for example, L.D.Landau and E.M.Lifshitz, *Mechanics* (Pergamon Press, Oxford, 1960).

$$\begin{aligned}
 \gamma &= \hbar \frac{\partial^2}{\partial L^2} \int_a^\infty [2m(E-U) - L^2/r^2]^{\frac{1}{2}} dr = \\
 &= -\frac{\partial}{\partial L} \int_a^\infty \frac{(L\hbar/r^2) dr}{[2m(E-U) - L^2/r^2]^{\frac{1}{2}}}.
 \end{aligned} \tag{90.13}$$

Making use of (90.11) we transform the expression for γ into the form

$$\gamma = \pm \frac{\hbar}{2} \frac{\partial \theta}{\partial L}.$$

If the value of B from (90.4) is substituted and the value found for γ is used, then the differential cross section takes the form

$$d\sigma = |f(\theta)|^2 d\Omega = \frac{L}{m^2 v^2 \sin \theta} \left| \frac{\partial L}{\partial \theta} \right| d\Omega. \tag{90.14}$$

Replacing the quantity L in (90.14) by its classical value, we obtain

$$d\sigma = \frac{\rho}{\sin \theta} \left| \frac{\partial \rho}{\partial \theta} \right| d\Omega. \tag{90.15}$$

Expression (90.15) represents the ordinary scattering formula given by classical mechanics.

We examine the limits of applicability of the formulae (90.15) for the scattering cross section. They can be established from the following obvious considerations.

One can speak of the motion of a particle in a trajectory in the case where the corresponding wavelength is small in comparison with the size of the system. In the given case the wavelength must be small in comparison with the size of the region in which a considerable interaction takes place. If the size of this region is denoted by R , then this requirement can be written in the form

$$\lambda \ll R, \tag{90.16}$$

where λ is the de Broglie wavelength.

Substituting the value of λ into formula (90.16) we find

$$R \gg \frac{2\pi\hbar}{mv}. \tag{90.17}$$

In order that the behaviour of the particle may be characterized by classi-

cal concepts, it is necessary that the quantum-mechanical uncertainties be small. In other words, it is necessary that the following relations be fulfilled:

$$\frac{\Delta\theta}{\theta} \ll 1, \quad \frac{\Delta\rho}{\rho} \ll 1, \quad (90.18)$$

where ρ is the classical impact parameter, and $\Delta\theta$ and $\Delta\rho$ are respectively the quantum-mechanical uncertainties for the scattering angle θ and the impact parameter ρ .

For the quantity $\Delta\theta$ one can write an expression valid in order of magnitude

$$\Delta\theta \sim \frac{\Delta p}{p}, \quad (90.19)$$

where Δp is the uncertainty in the transverse component of the momentum. Making use of the uncertainty relation for the coordinate and momentum

$$\Delta p \cdot \Delta\rho \sim \hbar$$

and eliminating the quantity Δp from (90.19), and then using (90.18), we obtain

$$\theta \gg \Delta\theta \sim \frac{\hbar}{\Delta\rho p} \gg \frac{\hbar}{\rho p}. \quad (90.20)$$

This condition assumes a considerably simpler form if the scattering angles are small. Namely, in this case the scattering angle θ can be found in a simple way. It is equal to the ratio of the value of the transverse momentum acquired by the scattered particle in traversing the field of the scatterer to the longitudinal momentum. The transverse momentum is equal to the force $U'(\rho)$ acting on the particle multiplied by the time τ for which this force acts: $\tau = \rho/v$. Thus the scattering angle θ is, in order of magnitude, equal to

$$\theta \approx |U'(\rho)| \frac{\rho}{vp}. \quad (90.21)$$

Or, substituting (90.21) into (90.20), we find the condition of applicability of the theory

$$|U'(\rho)|\rho^2 \gg \hbar v.$$

But if the derivative $U'(\rho)$ is replaced by $U(\rho)/\rho$, then the condition of applicability can be rewritten in the form

$$U(\rho) \gg \frac{\hbar v}{\rho}. \quad (90.22)$$

Comparing (90.22) with the condition of applicability of the Born approximation (84.12), we see that the conditions are opposite to each other. Thus these methods supplement each other to a considerable degree.

§91. The general theory of inelastic scattering and the absorption of particles

So far we have confined ourselves to consideration of the elastic scattering process. We now turn to the more general case where inelastic scattering is also possible.

Any process in which the internal state of the particles changes is said to be inelastic. Thus, for example, collisions accompanied by an excitation (for instance, an excitation of the atom or nucleus), by a decay or by the production of new particles, for example, are inelastic. Each of the possible processes is called a reaction channel. If the process is compatible with conservation laws, the channel is said to be open. In what follows we shall consider processes for which the inelastic and elastic reaction channels are open. We shall begin with a generalization of partial wave scattering theory. This will allow us to cover, at the same time, processes of elastic and inelastic scattering and absorption. For a formal description of any scattering process we shall surround the scattering centre by a fictitious sphere of sufficiently large radius R_0 .

Let us consider the character of the l th partial wave for $r > R_0$ in three cases:

- (1) at the origin there is no scattering centre,
- (2) at the origin there is a scattering centre at which the particle undergoes only elastic scattering,
- (3) at the origin there is a scattering centre at which the particle undergoes inelastic scattering.

In the first case the radial function of the l th partial wave can be written (see (36.10)) in the form of a superposition of two waves

$$R_l = a_l \frac{e^{i(kr - \frac{1}{2}\pi l)}}{2ikr} - a_l \frac{e^{-i(kr - \frac{1}{2}\pi l)}}{2ikr} = a_l \frac{\sin(kr - \frac{1}{2}\pi l)}{kr}.$$

The second term represents a converging wave, and the first term a diverging wave. Here we make use of asymptotic expressions, since by assumption R_0 is sufficiently large. The amplitudes and phases of the two waves are the same and the wave function R_l is the product of a real function and a constant factor. Hence the flux through a closed surface is equal to zero:

$$j_l = \frac{\hbar}{2mi} \int \left(\psi_l^* \frac{\partial \psi_l}{\partial r} - \psi_l \frac{\partial \psi_l^*}{\partial r} \right) r^2 d\Omega = 0,$$

where

$$\psi_l = P_l(\cos \theta) R_l(r).$$

In the second case the radial function of the l th partial wave is written, according to (86.2), in the form

$$R_l = B_l \frac{\sin(kr + \delta_l - \frac{1}{2}\pi l)}{kr} = F_l \frac{e^{2i\delta_l} e^{i(kr - \frac{1}{2}\pi l)} - e^{-i(kr - \frac{1}{2}\pi l)}}{2ikr}. \quad (91.1)$$

The amplitudes of the converging and diverging waves differ from each other by the phase factor $e^{2i\delta_l}$, where $|e^{2i\delta_l}| = 1$. In this case the total partial flux through the surface of the sphere is also equal to zero (the partial wave function depending on l is real). Hence it follows that the diverging and converging fluxes of the l th partial wave are equal to each other. The fact that the converging and diverging waves have different coefficients, $e^{2i\delta_l}$ and unity, does not contradict this equality, since $|e^{2i\delta_l}| = 1$.

In the third case, where the particles undergo inelastic scattering, it is impossible to write a general expression for the radial function taking into account all possible inelastic processes. We can, however, simplify the problem if we consider elastic scattering separately from all possible forms of inelastic scattering.

In this case we can write the following formal expression for the radial function of the l th partial wave describing the elastic scattering of a particle:

$$R_l = b_l \frac{S_l e^{i(kr - \frac{1}{2}\pi l)} - e^{-i(kr - \frac{1}{2}\pi l)}}{2ikr} \quad (91.2)$$

This expression is constructed according to the same principle as (91.1), but it takes into account the particular process, in which inelastic processes or absorption may exist along with elastic scattering. The coefficient S_l introduced is in magnitude less than one. This expresses the fact that in the presence of absorption or inelastic scattering the converging flux of elastically scattered particles is larger than the diverging flux. Then the wave function is written in the form

$$\psi = \sum_l b_l \frac{S_l e^{i(kr - \frac{1}{2}\pi l)} - e^{-i(kr - \frac{1}{2}\pi l)}}{2ikr} P_l(\cos \theta) = \sum_l \psi_l.$$

The coefficients b_l are again defined by the requirement that the wave func-

tion ψ be the same as (83.3). On carrying out calculations analogous to those for elastic scattering, we find the wave function in the form

$$\psi = \sum_{l=0}^{\infty} \frac{i^l}{2ikr} (2l+1) [S_l e^{i(kr - \frac{1}{2}\pi l)} - e^{-i(kr - \frac{1}{2}\pi l)}] P_l(\cos \theta). \quad (91.2')$$

It is easily shown that the flux of elastically scattered particles with given angular momentum through a sphere of radius $r \gg R_0$ is different from zero. Indeed, we have

$$\frac{\partial \psi_l}{\partial r} = \frac{i^l (2l+1) P_l(\cos \theta)}{2ikr} [ik S_l e^{i(kr - \frac{1}{2}\pi l)} + ik e^{-i(kr - \frac{1}{2}\pi l)}] P_l(\cos \theta).$$

Here only terms proportional to r^{-1} are retained in the expression for $\partial \psi_l / \partial r$. Terms proportional to r^{-2} are dropped, since we desire to find the flux through a sphere of large radius. We further calculate the total flux of particles through a sphere of radius $r \gg R_0$. It is equal to

$$j_l = \frac{i\hbar}{2m} r^2 \int \left(\psi_l^* \frac{\partial \psi_l}{\partial r} - \psi_l \frac{\partial \psi_l^*}{\partial r} \right) d\Omega.$$

Substituting the functions ψ_l and $\partial \psi_l / \partial r$ into the expression for the flux and taking into account the conditions of normalization of the Legendre polynomials $P_l(\cos \theta)$, we obtain

$$j_l = -\frac{\pi \hbar}{mk} (2l+1)(1-|S_l|^2). \quad (91.3)$$

Since $|S_l| \ll 1$, the flux is negative. This means that the total flux is directed inwards through the sphere.

It is easy to understand the meaning of this result: the flux of particles incident on the centre with angular momentum l turns out to be larger than the flux of elastically scattered particles. The particles undergo inelastic scattering or absorption, and the intensity of the beam of elastically scattered particles is reduced. It is clear that on dividing the flux j_l by the flux density of incident particles we find, by definition, the partial inelastic scattering cross section. Here inelastic scattering is understood to be all processes reducing the intensity of elastic scattering. Since the flux density of incident particles is equal to v , then for the l th partial inelastic scattering cross section we obtain

$$\sigma_{l\text{inel}} = \frac{\pi}{k^2} (2l+1)(1-|S_l|^2). \quad (91.4)$$

As for the elastic scattering amplitude, we can, without reproducing the calculations of §86, write for it the expression

$$f(\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1)(S_l - 1) P_l(\cos \theta), \quad (91.5)$$

since formula (91.1) differs from (91.2) by the substitution of S_l for $e^{2i\delta_l}$.

The set of complex quantities S_l defines the cross section for inelastic as well as elastic scattering. In particular, if $S_l = e^{2i\delta_l}$, where δ_l is real, then the inelastic scattering cross section reduces to zero, and the elastic scattering amplitude is the same as expression (86.11).

Besides the l th partial cross section for elastic and inelastic scattering processes one can also write the total cross sections for the processes.

The total inelastic scattering cross section is evidently equal to

$$\sigma_{\text{inel}} = \frac{\pi}{k^2} \sum_{l=0}^{\infty} (2l+1)(1 - |S_l|^2) = \sum_l \sigma_l, \quad (91.6)$$

and the total elastic scattering cross section is equal to

$$\sigma_{\text{el}} = \int |f(\theta)|^2 d\Omega = \frac{\pi}{k^2} \sum_{l=0}^{\infty} (2l+1) |1 - S_l|^2. \quad (91.7)$$

We now turn to the consideration of formula (91.6).

Each cross section σ_l can be pictured as a characteristic of the process of inelastic scattering or absorption of particles with angular momentum l . Since the quantity $|S_l|^2 < 1$, it can be stated that the partial cross section σ_l has the upper limit $\sigma_{l\text{max}} = \pi k^{-2}(2l+1)$.

The structure of formula (91.6) and the physical meaning of the coefficient $1 - |S_l|^2$ can easily be understood by means of the following reasoning based on the quasi-classical approximation.

The collision parameter of a particle can be (see (86.14)) written in the form

$$\rho_l = \frac{\hbar}{p} [l(l+1)]^{\frac{1}{2}}. \quad (91.8)$$

For large l we obtain

$$\rho_l = \frac{\hbar}{p} l.$$

The area of the annulus lying between two circles of radii ρ_l and ρ_{l+1} is equal to

$$2\pi\rho_l \frac{\hbar}{p} \approx \frac{\pi\hbar^2}{p^2} (2l+1).$$

The number of particles passing through this annulus oriented perpendicularly to the incident flux can easily be found. If the flux density of incident particles is equal to unity, then the number of particles crossing the ring is numerically equal to $\pi\hbar^2 p^{-2}(2l+1)$.

We introduce the so-called absorption coefficient ξ_l , which by definition represents the ratio of the number of absorbed particles incident on a given surface to the total number of particles incident on this surface. The number of particles absorbed by the surface of the annulus defined by radii ρ_l and ρ_{l+1} is defined by the expression $\pi\hbar^2 p^{-2}(2l+1)\xi_l$, and, correspondingly, the absorption cross section will have the form

$$\sigma_{l\text{inel}} = \frac{\pi}{k^2} (2l+1)\xi_l. \quad (91.9)$$

Comparing formulae (91.6) and (91.9) we see that

$$1 - |S_l|^2 = \xi_l, \quad (91.10)$$

i.e. the quantity $1 - |S_l|^2$ is the absorption coefficient.

Finally, we also obtain the formula relating elastic and inelastic scattering cross sections. It turns out that the following equality holds:

$$\frac{4\pi}{k} \text{Im } f(0) = \sigma_{\text{inel}} + \sigma_{\text{el}}. \quad (91.11)$$

To obtain this relation we shall calculate the sum of elastic and inelastic cross sections. By means of (91.6) and (91.7) we have

$$\sigma_{\text{inel}} + \sigma_{\text{el}} = \frac{\pi}{k^2} \sum_{l=0}^{\infty} (2l+1)(2 - S_l - S_l^*) = \frac{\pi}{k^2} \sum_{l=0}^{\infty} (2l+1)(2 - 2 \text{Re } S_l). \quad (91.12)$$

On the other hand, since the Legendre polynomials for $\theta = 0$ are equal to one, we have for the scattering amplitude

$$f(0) = \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1)(S_l - 1),$$

and the imaginary part of the scattering amplitude is equal to

$$\text{Im } f(0) = \frac{1}{2k} \sum_{l=0}^{\infty} (2l+1)(1 - \text{Re } S_l).$$

Comparing the expressions obtained, we see that eq. (91.11) is valid. Thus we have shown that the sum of inelastic and elastic scattering cross sections is proportional to the imaginary part of the scattering amplitude taken for the value of the angle $\theta = 0$. Formula (91.11) is called the optical theorem.

In conclusion we note that the absorption of particles can be described by introducing a complex potential $U = V_1 - iV_2$, where V_1 and V_2 are real functions. The imaginary part of the potential characterizes the absorption or emission of particles. As a matter of fact, in this case the Schrödinger equation has the form

$$i\hbar \frac{\partial \psi}{\partial t} = \left(-\frac{\hbar^2}{2m} \nabla^2 + V_1 - iV_2 \right) \psi. \quad (91.13)$$

Carrying out calculations analogous to those of §7, we obtain

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} - \frac{2V_2 \psi^* \psi}{\hbar} = 0, \quad (91.14)$$

where

$$\rho = \psi^* \psi, \quad \mathbf{j} = \frac{\hbar}{2mi} [\psi \nabla \psi^* - \psi^* \nabla \psi].$$

In the stationary case for V_2 equal to zero, $\nabla \cdot \mathbf{j} = 0$, which corresponds to the absence of absorption or emission of particles. If $V_2 \neq 0$, then we obtain

$$\nabla \cdot \mathbf{j} = 2V_2 \rho / \hbar.$$

Depending on the sign of V_2 this formula describes the absorption or emission of particles.

§92. The diffraction of fast neutrons by nuclei

The study of the interaction of fast neutrons with nuclei shows that in the region of neutron energies above a few tens of MeV for light nuclei and a few hundreds of MeV for heavy nuclei a very intense capture of neutrons takes place.

The intense absorption of neutrons is also accompanied by their elastic scattering. In describing the strong absorption of fast neutrons the following optical analogy turns out to be very useful. The nucleus behaves with respect to the neutrons as a perfectly absorbing (black) sphere on which a light wave is incident. The absorption of the light wave by the black sphere is accompanied by its perturbation in the region of space near the absorber. This means that in addition to absorption, light scattering occurs. Analogously to this the absorption of neutrons by a nucleus will perturb their wave function and the neutrons will undergo elastic scattering.

To calculate the neutron elastic scattering cross section we shall make use of the analogy with optical phenomena. In §36 of Part IV we have seen that diffraction phenomena occur when the wavelength of the light is less than the radius of the scattering sphere. In this case the intensity of light scattered by a black sphere of radius R into solid angle $d\Omega$ is given by expression (36.13) of Part IV

$$dI = \frac{I}{\pi} \left| \frac{J_1(kR\theta)}{\theta} \right|^2 d\Omega, \quad (92.1)$$

where θ is the angle between the direction of motion of the scattered light and the initial direction of its incidence, I is the total intensity of light incident on the screen, and J_1 is the Bessel function of first order.

Simple estimation shows that the wavelength of neutrons of an energy of the order of 1 MeV is smaller by a factor of several hundred than the nuclear size. Therefore the optical formula (92.1) can be applied to the scattering of neutrons by the absorbing nucleus. To obtain the neutron differential scattering cross section the flux of neutrons scattered into angle $d\Omega$ must be divided by the incident neutron flux density $I/\pi R^2$. We then have

$$d\sigma = R^2 \left| \frac{J_1(kR\theta)}{\theta} \right|^2 d\Omega. \quad (92.2)$$

This expression, of course, can also be obtained from the general formula (91.5).

From the condition of 'blackness' of the nucleus it follows that the

absorption coefficient ξ_l is equal to one for those l for which $\rho < R$ and $\xi_l = 0$, if $\rho_l > R$. Since $\rho_l \sim \hbar l / p$ (see §91), then

$$S_l = \begin{cases} 0 & l < pR/\hbar = kR, \\ 1 & l > pR/\hbar = kR, \end{cases}$$

where $kR \gg 1$. Substituting these values of S_l into (91.5) we find the elastic scattering amplitude

$$f(\vartheta) = -\frac{1}{2ik} \sum_{l=0}^{kR} (2l+1) P_l(\cos \vartheta).$$

The dominant part in the sum is played by terms with large l . Hence we can disregard unity in comparison with $2l$, and use for the Legendre polynomial $P_l(\cos \vartheta)$ the approximate expression valid for $\vartheta \ll 1^*$.

$$P_l(\cos \vartheta) = J_0[(l + \frac{1}{2})\vartheta] \cong J_0(l\vartheta)$$

and pass from a summation over l to an integration

$$f(\vartheta) = \frac{i}{k} \int_0^{kR} l J_0(l\vartheta) dl = \frac{iR}{\vartheta} J_1(kR\vartheta).$$

From which we immediately obtain expression (92.2) for the cross section.

Let us consider the dependence of the differential cross section (92.2) on the scattering angle θ in more detail. The differential cross section does not depend on the azimuthal angle. Evidently we have

$$d\sigma = 2\pi R^2 \left| \frac{J_1(kR\theta)}{\theta} \right|^2 \sin \theta d\theta. \quad (92.3)$$

For small angles $kR\theta < 1$, we expand the Bessel function in a series and find $J_1(kR\theta) \approx \frac{1}{2}kR\theta$. Consequently, for small angles the cross section assumes the form

$$d\sigma = \frac{1}{4}k^2 R^4 d\Omega. \quad (92.4)$$

This is independent of the scattering angle θ .

* See, for example, L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

For larger angles up to values lying in the interval $1 \gg \theta \gg 1/kR$ one can write the asymptotic expression for the Bessel function

$$J_1(kR\theta) \approx \left(\frac{2}{\pi kR\theta}\right)^2 \sin(kR\theta - \frac{1}{4}\pi).$$

In this range of angles the cross section, which oscillates, decreases rapidly with increasing θ . The value of the cross section at the maxima decreases in proportion to θ^{-3} .

Thus the cross section has a sharp maximum for scattering at an angle $\theta \approx 0$, i.e. for forward scattering in directions close to the direction of the incident beam.

The total scattering cross section σ can be found by integrating (92.3) over the entire solid angle,

$$\sigma = 2\pi R^2 \int \frac{J_1^2}{\theta^2} \sin \theta \, d\theta.$$

In view of the rapid convergence of the integral the contribution to it by large values of θ is small, and the upper limit of the integral can approximately be replaced by infinity. Then, making use of the formula

$$\int_0^\infty \frac{J_1^2(x)}{x} \, dx = \frac{1}{2},$$

we find finally

$$\sigma = \pi R^2. \quad (92.5)$$

The total cross section for the scattering of neutrons with $\lambda \leq R$ is the same as the geometric cross section of the nucleus.

Let us also define the total cross section for the absorption of neutrons by a nucleus. Making use of the expression for S_l and substituting it into formula (91.6) we obtain

$$\sigma_{\text{inel}} = \frac{\pi}{k^2} \sum_{l=0}^{kR} (2l+1) = \pi R^2. \quad (92.6)$$

Consequently, the cross section for the absorption of neutrons by a black nucleus is also the same as the geometric cross section of the nucleus.

From relations (92.5) and (92.6) it follows that the total cross section for the interaction of neutrons with a nucleus is equal to twice the geometric cross section of the nucleus,

$$\sigma_{\text{incl}} + \sigma_{\text{el}} = 2\pi R^2. \quad (92.7)$$

By analogous methods one can also calculate the cross sections for scattering by nuclei which only partially absorb the neutrons incident on them, as well as the diffraction of charged particles by nuclei*.

§93. The scattering of slow particles. The threshold approximation

We shall apply the partial wave scattering formulae already obtained to find the cross sections for the elastic and inelastic scattering of slow particles. As in §86, we shall understand slow particles to be those whose de Broglie wavelength λ is large compared with the size of the region of interaction. We shall restrict ourselves to the case where the interaction energy decreases sufficiently rapidly with increasing distance, that an effective radius R of the interaction sphere can be introduced.

As we have seen in §86, only s-wave scattering is significant at small energies. The radial part of the wave function corresponding to the angular momentum $l = 0$ satisfies eq. (35.8)

$$-\frac{\hbar^2}{2m} \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial R_0}{\partial r} \right) + U(r) R_0 = E R_0.$$

Introducing the wave number $k = (2mE/\hbar^2)^{1/2}$, the above equation can be rewritten in the form

$$R_0'' + \frac{2}{r} R_0' + k^2 R_0 - \frac{2m}{\hbar^2} U(r) R_0 = 0. \quad (93.1)$$

For $r > a$, the potential energy is equal to zero outside the interaction region and the equation for the function R_0 assumes the form

$$R_0'' + \frac{2}{r} R_0' + k^2 R_0 = 0. \quad (93.2)$$

It is clear that the potential energy does not reduce to zero sharply at a certain limit, but changes over a transitional region according to a complex and usually unknown law.

Therefore at first sight the finding of the wave function over all space seems to be an extremely complex problem. In reality, however, this is not so.

* For more details see A.I. Akhiezer and I. Ya. Pomeranchuk, *Nekotorye voprosy teorii yadra* (Some problems in nuclear theory) (Gostekhizdat, Moscow, 1950).

It turns out that use can be made of the large value of λ or, what is the same, of the small k , essentially to simplify the problem. Namely, in the region $r < a$, i.e. in the region of effective interaction, the term k^2 can be neglected since it is small in comparison with $2m\hbar^{-2}U(r)$. Then we have

$$R_0'' + \frac{2}{r} R_0' - \frac{2m}{\hbar^2} U(r) R_0 = 0. \quad (93.3)$$

The solution of eq. (93.3) for a given function $U(r)$ can be written in the form $R_0(r, c_1, c_2)$, where c_1 and c_2 are two arbitrary constants. Since (93.3) does not involve the quantity k , we shall assume the wave function to be independent of k in the region $r < a$.

The solution of eq. (93.2) is

$$R_0 = \frac{1}{2ikr} (c_3 e^{ikr} + c_4 e^{-ikr}). \quad (93.4)$$

Here c_3 and c_4 are two constants of integration which do not depend on r but, generally speaking, are functions of k .

The equation for the wave function cannot be written for the transitional region, since the behaviour of the potential energy here is unknown. However, the width of the intermediate region is small in comparison with the size of the interaction region and is very small in comparison with the wavelength λ .

Nevertheless, a substantial change of the wave function takes place over a distance λ . Hence one can disregard the change of the wave function in the transition region and replace it by a sharp boundary at $r = a$. At this surface the two solutions must join smoothly.

It is clear, however, that two functions, one of which depends on k as a parameter and the other of which does not depend on k at all, can only be joined when in the neighbourhood of the boundary of the region the function (93.4) also becomes independent of k .

For $r \sim a$ the quantity ka is small by definition. Hence expanding the exponentials (93.4) in a series in powers of kr and retaining the first two terms of the expansion, we obtain

$$R_0 = \frac{c_3(k)(1+ikr) + c_4(k)(1-ikr)}{2ikr}. \quad (93.5)$$

This expression will not depend on k when the following relations are fulfilled

$$\begin{aligned} c_4(k) + c_3(k) &= 2ika_2, \\ c_3(k) - c_4(k) &= 2a_1, \end{aligned} \quad (93.6)$$

where a_1 and a_2 are constant quantities independent of k .

Solving the system of equations (93.6), we find for c_3 and c_4

$$\begin{aligned} c_4 &= -a_1 + ika_2, \\ c_3 &= ika_2 + a_1. \end{aligned} \quad (93.7)$$

Comparing expression (93.4) with (93.3), we find the quantity S_0 to have the form

$$S_0 = -\frac{c_3}{c_4} = -\frac{a_1 + ika_2}{ika_2 - a_1}. \quad (93.8)$$

Expanding in a series in small values of k , we obtain

$$S_0 = 1 + 2ika_2/a_1.$$

Making use of formulae (91.6) and (91.7) we find the expression for the cross sections for elastic and inelastic processes:

$$\begin{aligned} \sigma_{el} &= \frac{\pi}{k^2} |1 - S_0|^2 = 4\pi \left| \frac{a_2}{a_1} \right|^2, \\ \sigma_{inel} &= \frac{\pi}{k^2} (1 - |S_0|^2) = \frac{4\pi}{k} \operatorname{Im} \frac{a_2}{a_1}. \end{aligned} \quad (93.9)$$

From these formulae it follows that the elastic scattering cross section in the case considered does not depend on the energy of the scattered particle. The inelastic scattering cross section is inversely proportional to the wave number k , i.e. inversely proportional to the velocity v of the particle.

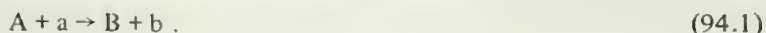
The method which we have used of neglecting the width of the transition zone and the replacement of eq. (93.1) by eq. (93.2) in the internal region is of a very general character and is called the threshold approximation. This approximation can be applied successfully in all cases where the wavelength can be considered to be large in comparison with the width of the transition region.

We shall encounter the use of the threshold approximation later.

§94. The Breit-Wigner formula

In the preceding sections we have considered the laws of elastic scattering

as well as of the absorption of particles. We now turn to the study of some phenomena occurring in nuclear reactions of the type



Here A and B are the initial and final nuclei, a is the incident particle and b is the particle emerging as a result of the reaction. In order to avoid complications associated with the effect of the nuclear electric field, we shall confine ourselves to the case where the incident and outgoing particles are neutrons.

The study of reactions caused by neutrons of relatively small energies showed that the cross sections for the reactions as a function of the energy of the incident neutrons display maxima at definite energy values. The phenomenon is of a pronounced resonant character and the maxima correspond to very narrow neutron energy intervals.

To account for the resonant character of nuclear reactions Bohr proposed the following general scheme of nuclear reactions. Neutron a , penetrating the nucleus, strongly interacts with the nuclear particles and transfers its excess energy to them. This latter energy is evidently equal to the sum of its kinetic energy and the binding energy of the particle in the nucleus U_0 . The energy brought by the neutron is rapidly distributed among all the nucleons in the nucleus, since they interact strongly with each other. As a result a new, so-called compound nucleus C arises from nucleus A and the neutron. The compound nucleus is not a stable system, since its energy is higher than the energy of the normal state by an amount $E + U_0$. After a certain lapse of a time the compound nucleus will make a transition to the normal state. At small excitation energies this transition may proceed in one of two ways.

First, as a result of a fluctuation all the excitation energy may be concentrated on one of the nuclear particles. This particle (for simplicity of the argument, a neutron) then has the possibility of escaping from the nucleus, with energy E . Evidently this mode of reaction corresponds to the elastic scattering of the neutron by the nucleus.

Secondly, only part of the excitation energy may be carried away by the escaping neutron. The rest of the excitation energy is emitted by the system of nuclear particles in the form of a γ -quantum. In this case inelastic scattering of the neutron occurs. A particular case of this reaction is the radiative capture of the neutron, in which the entire excitation energy is carried away by a γ -quantum and the neutron remains in the nucleus.

For the cross sections for the elastic and inelastic scattering use can be made of formulae (91.6) and (91.7). We shall restrict ourselves to the case of slow neutrons, described by an s -wave, and shall consider the nucleus to be a sphere of radius R . Although the nucleus cannot be considered to have sharp

geometric bounds, its diffuseness is very small in comparison with the wavelength of the incident neutron $\lambda \gg R$.

The neutron inside the nucleus must be in a state to which there corresponds a wavelength λ_{int} . At the surface $r = R$ the wave functions describing the neutron outside and inside the nucleus must join for which the following conditions are to be fulfilled:

$$\psi = \psi_{\text{int}}, \quad \frac{d\psi}{dr} = \frac{d\psi_{\text{int}}}{dr}.$$

It follows from the second condition that the orders of magnitude of the amplitudes of the wave functions of the external and internal motions are in the ratio $\sim \lambda_{\text{int}}/\lambda$. This means that the probability for the particle to get inside the nucleus is $\sim (\lambda_{\text{int}}/\lambda)^2$, i.e. is very small.

The corresponding energy is determined by the value of the normal derivative of the wave function at the surface of the nucleus.

We denote by $f(E)$ the quantity

$$f(E) = R \left(\frac{d(r\psi)/dr}{r\psi} \right)_{r=R}. \quad (94.2)$$

The quantity $f(E)$ is related directly to the normal derivative $(d\psi/dr)_{r=R}$ and depends on the neutron energy E . The quantity S_0 , defining the cross section for the elastic and inelastic scattering of the s-wave, can easily be expressed in terms of f .

Substituting the value of ψ from (91.2) into (94.2), we find

$$f = -i \frac{kR e^{-ikR} + kRS_0 e^{ikR}}{e^{-ikR} - S_0 e^{ikR}}.$$

Hence it follows that

$$S_0 = -e^{-2ikR} \frac{kR - if(E)}{kR + if(E)}. \quad (94.3)$$

Since $f(E)$ is, generally speaking, a complex quantity, one can write

$$f(E) = f_1(E) - if_2(E), \quad (94.4)$$

where $f_1(E)$ and $f_2(E)$ are real functions. Since $|S_0|$ is always ≤ 1 , then the function $f_2(E) \geq 0$.

Taking into account (94.4) we have for S_0

$$S_0 = -e^{-2ikR} \frac{kR - if_1(E) - f_2(E)}{kR + if_1(E) + f_2(E)}. \quad (94.5)$$

Substituting this value of S_0 into (91.6), we find

$$\sigma_{\text{inel}} = \frac{\pi}{k^2} (1 - |S_0|^2) = \frac{4\pi}{k^2} \frac{kRf_2}{(kR + f_2)^2 + f_1^2}. \quad (94.6)$$

Analogously from (91.7) it follows that

$$\begin{aligned} \sigma_{\text{el}} &= \frac{\pi}{k^2} |1 - S_0|^2 = \frac{\pi}{k^2} \left| 1 + e^{-2ikR} \frac{kR - if_1 - f_2}{kR + if_1 + f_2} \right|^2 = \\ &= \frac{4\pi}{k^2} \left| e^{-ikR} \frac{kR \cos kR - f_1 \sin kR + if_2 \sin kR}{kR + if_1 + f_2} \right|^2 = \\ &= \frac{4\pi}{k^2} \left| \frac{kR}{i(kR + f_2) - f_1} + e^{ikR} \sin kR \right|^2. \end{aligned} \quad (94.7)$$

Let us first discuss the formula for σ_{inel} . Since $f_2 > 0$, the cross section has a maximum for $f_1(E_0) = 0$. When the neutron has an energy equal to E_0 it has a relatively high probability of penetrating the nucleus. Accordingly the energy E_0 corresponds to a resonance value of the energy of the nucleus. Near the resonance energy we can expand the function $f_1(E)$ in a series in powers of $E - E_0$ and restrict ourselves to the first term of the expansion

$$f_1(E) = f'(E_0)(E - E_0).$$

It can be shown* that the quantity $f'(E_0) < 0$. We introduce the notation

$$\Gamma_e = -\frac{2kR}{f'(E_0)}; \quad \Gamma_r = -\frac{2f_2}{f'(E_0)}; \quad \Gamma = \Gamma_e + \Gamma_r. \quad (94.8)$$

Then we find

$$\sigma_{\text{inel}} = \frac{\pi}{k^2} \frac{\Gamma_e \Gamma_r}{(E - E_0)^2 + \frac{1}{4}\Gamma^2}. \quad (94.9)$$

In formula (94.7) for σ_{el} for $E \approx E_0$ the first term is usually large in comparison with the second, and one can write

$$(\sigma_e)_{\text{el}} \approx \frac{\pi}{k^2} \frac{\Gamma_e^2}{(E - E_0)^2 + \frac{1}{4}\Gamma^2}. \quad (94.10)$$

The formulae for the cross sections for elastic and inelastic scattering of slow neutrons are called the Breit-Wigner formulae. To explain the physical meaning of the quantities Γ_e , Γ_r and Γ which have been introduced it is useful to

* See A.I. Akhiezer and I.Ya. Pomeranchuk, *Nekotorye voprosy teorii yadra (Some problems in nuclear theory)* (Gostekhizdat, Moscow, 1950).

compare the Breit-Wigner formulae with the dispersion formulae of the theory of light scattering (§ 108). We see that the general structure of the formulae is the same. This is quite natural, since the Breit-Wigner formulae could be obtained by considering the reaction as the transition of the system (nucleus + neutron) from the initial into the final state via a compound nucleus as an intermediate state (i.e. according to the same scheme as in the scattering of photons). Direct application of perturbation theory leads to the Breit-Wigner formulae. However this way of obtaining the Breit-Wigner formulae cannot be substantiated, since the perturbation of the state of the neutron is not weak.

Nevertheless, such an obvious although not rigorous calculation shows that the quantities Γ_e and Γ_r characterize transition probabilities. Namely, Γ_e is proportional to the matrix element of the transition of the system from the intermediate state (nucleus C) to the final state (nucleus A and the neutron with energy E). Hence the quantity Γ_e , which is called the partial width of the resonance level E corresponding to elastic scattering, determines the probability of decay of the nucleus C with elastic scattering of the neutron. The quantity Γ_r is called the partial width of the resonance level with respect to the reaction. It determines the probability of decay of the nucleus with inelastic scattering and neutron capture. In the case of slow neutrons the probability of inelastic scattering is small and the reaction reduces to the capture of the neutron. Finally, Γ determines the total probability of decay of nucleus C. It is equal to the energy half-width of the resonance maximum of the cross section.

Fig. V.27 shows the energy dependence of the cross sections for the elastic scattering and radiative capture of slow neutrons. We have introduced the Breit-Wigner formulae in the particular case where the energy of the neutron is close to one of the resonance levels E_0 of the nucleus. They can be generalized to the case of many levels. They may also take into account the spin states of the nucleus and of light particles. Finally, the Breit-Wigner formulae can be generalized to the case of charged particles and particles with an angular momentum. We discuss certain properties of the widths of the resonance levels of a nucleus. The reaction width Γ_r for slow neutrons reduces to the radiative capture width Γ_γ , since no inelastic scattering occurs at small energies. The value of Γ_γ amounts to about 10^{-1} eV and does not depend on the velocity of the neutron. The width $\Gamma_e \sim k \sim v$, where v is the velocity of the neutron, and at small energies for heavy and medium nuclei $\Gamma_e \ll \Gamma_r$. This means that neutron capture predominates over elastic scattering. In the case of light nuclei the situation is the reverse: resonance scattering predominates over capture.

The Bohr concept of the formation of a compound nucleus is valid for

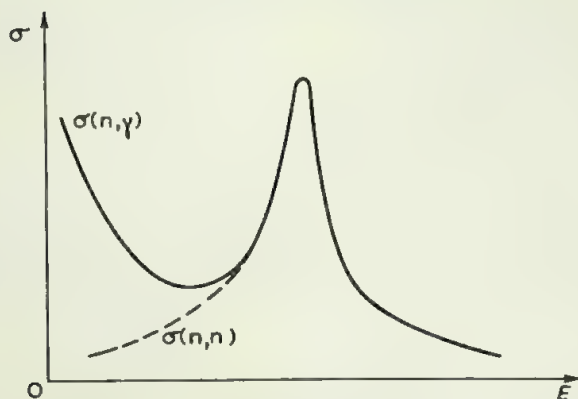


Fig. V.27

nuclear reactions proceeding at not too large energies. As the energy of the incident particles increases their cross section for scattering by individual nuclear nucleons decreases sharply. Hence at energies $E > 50\text{--}100$ MeV the interaction of the particles with the nucleus reduces to an interaction with an individual nucleon. The Breit-Wigner formulae turn out to be no longer applicable.

§95. The scattering matrix (*S*-matrix)

The mathematical technique of scattering theory described above is associated with an explicit form of the interaction potential distribution over all space. However, in a number of important cases there is no potential energy (independent of the velocity). Hence in modern scattering theory an important role is played by a more general statement of the problem. Let the wave function $\psi_a(t \rightarrow -\infty)$ of a system of particles be given in the initial state before the interaction. The general problem of scattering theory is to find the wave function of the system a long time after the interaction, $\psi(t \rightarrow \infty)$. The wave function $\psi(t \rightarrow \infty)$ can be expressed in terms of the initial function $\psi_a(t \rightarrow -\infty)$ by means of the operator $\hat{V}(t, t_0)$ introduced in §49 and describing the development of the wave function in time. By the scattering matrix S we shall mean the limiting expression of the operator $\hat{V}(t, t_0)$ (in the interaction representation) describing the development of the process in time (see §49).

$$\hat{S} = \lim_{\substack{t_0 \rightarrow -\infty \\ t \rightarrow \infty}} \hat{V}(t, t_0). \quad (95.1)$$

Thus the scattering matrix \hat{S} carries out the transformation of the initial state $\psi_a(-\infty)$ into the final state $\psi(\infty)$,

$$\psi(\infty) = \hat{S}\psi_a(-\infty). \quad (95.2)$$

The index a denotes the complete set of quantum numbers defining the state of the system before scattering. It is assumed that the particles in the initial as well as in final state are separated by sufficiently large distances from each other that the interactions between them need not be taken into account (the so-called adiabatic hypothesis).

We expand the function ψ in a series in terms of a certain complete system of functions ψ_b , where b denotes the corresponding set of quantum numbers

$$\psi = \sum_b c_b \psi_b.$$

Here the symbol \sum_b denotes summation over a discrete sequence of quantum numbers and integration over quantum numbers changing continuously.

As follows from (95.2), the expansion coefficients c_b are expressed in terms of the matrix elements of the operator \hat{S}

$$c_b = (\psi_b, \psi) = (\psi_b, \hat{S}\psi_a) = \langle b | \hat{S} | a \rangle = S_{ba}. \quad (95.3)$$

The square of the amplitude c_b gives the total probability of the transition of the system in scattering from state a into state b

$$W'_{ba} = |S_{ba}|^2. \quad (95.4)$$

Thus the matrix elements of the operator \hat{S} are directly related to the corresponding transition probabilities.

Since the operator $\hat{V}(t, t_0)$ is unitary (see §96), its limiting value is also unitary, i.e. for the operator \hat{S} we can write

$$\hat{S}\hat{S}^\dagger = \hat{S}^\dagger\hat{S} = \hat{I}, \quad (95.5)$$

where \hat{I} denotes the unit operator.

Taking the diagonal matrix elements of one of the relations (95.5), we obtain the obvious result

$$\sum_b S_{ab}^\dagger S_{ba} = \sum_b |S_{ba}|^2 = 1, \quad (95.6)$$

i.e. the sum of the probabilities of all possible transitions is equal to one. Using relation (95.4), one can express the cross section for the process in terms of the matrix elements of the operator \hat{S} . However, it is first necessary to obtain the expression for the transition probability per unit time.

We assume that the initial state ψ_a is characterized by a definite energy value E_a . The total energy of the system is conserved in time. Hence the matrix S_{ba} can be written in the form

$$S_{ba} = S_{ba}^E \delta(E_a - E_b).$$

The matrix S_{ba}^E is said to be given on the energy surface. Then the total transition probability (95.4) is written in the form

$$W'_{ba} = |S_{ba}^E|^2 \delta^2(E_a - E_b). \quad (95.7)$$

This probability is proportional to the square of the δ -function. We write one of the δ -functions in the form (see Vol. I, Appendix III)

$$\delta(E_a - E_b) = \lim_{T \rightarrow \infty} \int_{-\frac{1}{2}T}^{\frac{1}{2}T} \frac{1}{2\pi\hbar} \exp((i/\hbar)(E_b - E_a)t) dt.$$

Substituting this expression into (95.7) and integrating the transition probability over the energy of the final state, we obtain the transition probability in time T

$$\bar{W}_{ba} = \int W'_{ba} dE_b = \frac{1}{2\pi\hbar} |S_{ba}^E|^2 T. \quad (95.8)$$

Hence we find for the transition probability per unit time

$$W = \frac{1}{2\pi\hbar} |S_{ba}^E|^2. \quad (95.9)$$

To find the cross section for the process we have to divide the transition probability by the incident particle flux density.

In the initial state there are two particles. As usual we consider the scattering process in the centre-of-mass system. The wave function of the initial state ψ_a describes states with given energy of relative motion E_a and a direction of the momentum of relative motion $\mathbf{n}_a = \mathbf{p}_a/p_a$, and is normalized by the condition

$$\int \psi_{E_a \mathbf{n}}^* \psi_{E'_a \mathbf{n}'} dV = \delta(E_a - E'_a) \delta(\mathbf{n}_a - \mathbf{n}'_a) = p_a^2 \frac{dp}{dE} \delta(\mathbf{p}_a - \mathbf{p}'_a). \quad (95.10)$$

Then

$$\psi_{E_a \mathbf{n}} = |E_a, \mathbf{n}\rangle = p_a \left(\frac{dp}{dE} \right)^{\frac{1}{2}} \psi_{\mathbf{p}_a} = \frac{p_a}{v_a^{\frac{1}{2}}} |\mathbf{p}_a\rangle. \quad (95.11)$$

The incident particle flux density is equal to

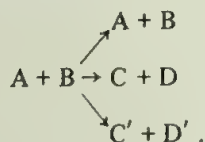
$$\mathbf{j}_0 = \frac{p_a^2}{(2\pi\hbar)^3} \mathbf{n}_a. \quad (95.12)$$

As always when we deal with a continuously changing quantity we have to introduce the differential transition probability dW_{ba} and, consequently, the differential cross section $d\sigma_{ba}$. Denoting the solid angle interval in which the vector \mathbf{n}_b lies by $d\Omega_b$, we obtain from (95.9) and (95.12)

$$d\sigma_{ba} = \frac{4\pi^2}{k_a^2} |\langle b, E, \mathbf{n}_b | S^E | a, E, \mathbf{n}_a \rangle|^2 d\Omega_b, \quad (95.13)$$

where $k_a = \hbar^{-1} p_a$.

Let us now consider the case where elastic and different forms of inelastic scattering may occur as a result of the interaction of two particles, i.e.



We shall call each form of transformation a reaction channel. Formula (95.13) for $b \neq a$ corresponds to an inelastic reaction channel. The cross section taking into account elastic and inelastic channels can be written in the form

$$d\sigma_{ba} = \frac{4\pi^2}{k_b^2} |\langle b, E, \mathbf{n}_b | S^E - I | a, E, \mathbf{n}_a \rangle|^2 d\Omega_b, \quad (95.14)$$

where I is the unit matrix. Since in the matrix I only diagonal elements are different from zero, for $b \neq a$ the cross section (95.14) is the same as (95.13).

Expressions (95.13) and (95.14) can be written in a form analogous to (86.12) if the initial state $|a, E, \mathbf{n}_a\rangle$ is expanded in terms of partial waves

$$|a, E, \mathbf{n}_a\rangle = |a, E, l, m\rangle \langle l, m | \mathbf{n}_a \rangle. \quad (95.15)$$

The transformation functions $\langle l, m | \mathbf{n}_a \rangle$ were found in §48:

$$\langle l, m | \mathbf{n}_a \rangle = Y_{lm}^*(\mathbf{n}_a). \quad (95.16)$$

Choosing the z -axis to be along the direction of the vector \mathbf{n}_a , we obtain

$$Y_{lm}(\mathbf{n}_a) = Y_{lm}(0) = \left(\frac{2l+1}{4\pi} \right)^{\frac{1}{2}} \delta_{m0}. \quad (95.17)$$

In substituting expressions (95.16), (95.17) into (95.13) and (95.14) matrix elements of the following form arise:

$$\langle b, E, \mathbf{n}_b | S^E | a, E, l, 0 \rangle.$$

In motion in a central field angular momentum is conserved. Hence the S -matrix is diagonal with respect to the quantum numbers l, m , and one can write

$$\begin{aligned} \langle b, E, \mathbf{n}_b | S^E | a, E, l, 0 \rangle &= \langle \mathbf{n}_b | l, 0 \rangle \langle b, E, l, 0 | S^E | a, E, l, 0 \rangle = \\ &= Y_{l0}(\mathbf{n}_b) \langle b, E, l, 0 | S^E | a, E, l, 0 \rangle = P_l(\cos \theta_b) \left(\frac{2l+1}{4\pi} \right)^{\frac{1}{2}} S_{ba}^l. \end{aligned} \quad (95.18)$$

Correspondingly for the differential cross section for scattering into solid angle $d\Omega_b$ we obtain

$$d\sigma_{ba} = \frac{1}{4k_a^2} \left| \sum_l (2l+1) (S_{ba}^l - \delta_{ba}) P_l(\cos \theta) \right|^2 d\Omega_b. \quad (95.19)$$

Integrating this expression over all directions of the vector \mathbf{n}_b , we obtain the cross section for the scattering $a \rightarrow b$

$$\sigma_{ba} = \frac{\pi}{k_a^2} \sum_l (2l+1) |S_{ba}^l - \delta_{ba}|^2. \quad (95.20)$$

From this formula it follows that the total elastic scattering cross section has the form

$$\sigma_{aa} = \frac{\pi}{k_a^2} \sum_l (2l+1) |S_{aa}^l - 1|^2. \quad (95.21)$$

We can also write down the expression for the total cross section for all inelastic processes σ_{inel} , which is obtained by summing σ_{ba} over all channels $b \neq a$

$$\sigma_{\text{inel}} = \sum_{b \neq a} \sigma_{ba} = \frac{\pi}{k_a^2} \sum_{b \neq a} \sum_l (2l+1) |S_{ba}^l|^2.$$

This expression can be transformed by making use of the unitarity of the S -matrix. Namely, we have

$$\sum_{b \neq a} |S_{ba}^l|^2 = 1 - |S_{aa}^l|^2. \quad (95.22)$$

Correspondingly for σ_{inel} we obtain

$$\sigma_{\text{inel}} = \frac{\pi}{k_a^2} \sum_l (2l+1)(1 - |S_{aa}^l|^2). \quad (95.23)$$

Formulae (95.23) and (95.21) are the same as formulae (91.6) and (91.7) of partial wave scattering theory. We see that the quantities S_l introduced in §91 are the diagonal matrix elements of the scattering matrix S . If inelastic processes are impossible, i.e. $S_{ba}^l = 0$ for $b \neq a$, then from unitarity relation (95.22) it follows that $|S_{aa}^l|^2 = 1$, i.e. that

$$S_{aa}^l = e^{2i\delta_l}. \quad (95.24)$$

Then expressions (95.19)–(95.21) are the same as the expressions for the elastic scattering cross section obtained in §86.

From relations (95.19), (95.21) and (95.22) it follows that the cross section for a process is determined by the matrix elements of the operator \hat{F} , $i\hat{F} = \hat{S} - \hat{I}$ (the factor i is introduced for convenience). The unitarity of the S -matrix leads to the following relation:

$$\hat{S}^\dagger \hat{S} = (\hat{I} - i\hat{F}^\dagger)(\hat{I} + i\hat{F}) = \hat{I},$$

or

$$-i\hat{F} + i\hat{F}^\dagger = \hat{F}^\dagger \hat{F}.$$

Taking the matrix elements of the left-hand and right-hand sides of this relation with respect to the wave functions (95.11), we obtain

$$\hat{F}_{ba} - \hat{F}_{ba}^\dagger = i \sum_c \hat{F}_{bc}^\dagger \hat{F}_{cb}, \quad (95.25)$$

where Σ_c denotes the summation over the discrete and integration over the continuous states of the system of two particles after collision. In fact, the system (95.25) is a system of integral equations expressing the property of unitarity of the S -matrix.

The system of equations (95.25) is substantially simplified if only elastic

scattering is possible. As is seen from comparison of (95.19), (86.11) and (86.12), in this case the matrix elements of the operator \hat{F} are to within a factor the same as the elastic scattering amplitude $f(\mathbf{n}', \mathbf{n})$

$$\frac{2\pi}{k} \hat{F}_{\mathbf{n}', \mathbf{n}} = f(\mathbf{n}', \mathbf{n}), \quad (95.26)$$

where \mathbf{n} and \mathbf{n}' are the unit vectors characterizing the direction of the momentum vector of the relative motion of the incident and scattered particles. From (95.25) we obtain

$$f(\mathbf{n}', \mathbf{n}) - f^*(\mathbf{n}, \mathbf{n}') = \frac{ik}{2\pi} \int f^*(\mathbf{n}'', \mathbf{n}') f(\mathbf{n}'', \mathbf{n}) d\Omega'' . \quad (95.27)$$

Relation (95.27) expresses the condition of unitarity for elastic scattering. For scattering in a central field the amplitude f depends only on the angle ϑ between the vectors \mathbf{n} and \mathbf{n}' , and relation (95.27) can be rewritten in the form

$$\text{Im } f(\mathbf{n}', \mathbf{n}) = \frac{k}{4\pi} \int f^*(\mathbf{n}'', \mathbf{n}') f(\mathbf{n}'', \mathbf{n}) d\Omega'' . \quad (95.28)$$

For $\mathbf{n} = \mathbf{n}'$ we obtain the relation connecting the imaginary part of the amplitude of the scattering at zero angle with the total cross section (optical theorem; see §91).

We note that eq. (95.28) makes it possible, in principle, to find the scattering amplitude if its modulus, which is defined by the scattering law, is known. Setting

$$f(\vartheta) = \left(\frac{d\sigma}{d\Omega} \right)^{\frac{1}{2}} e^{i\alpha(\vartheta)}$$

and substituting this expression into (95.28), we obtain the integral equation for the phase $\alpha(\vartheta)$. Thus knowing the scattering cross section $d\sigma/d\Omega$ we can, in principle, also determine the scattering amplitude $f(\vartheta)$. We note, however, that eq. (95.28) does not change under the replacement $\alpha \rightarrow \pi - \alpha$, i.e. it determines the scattering amplitude with an accuracy to within the transformation $f(\vartheta) \rightarrow -f^*(\vartheta)$.

Let us consider the effect of this uncertainty on the values of the scattering phase shifts. For this we calculate the integral

$$\begin{aligned} \int |f(\theta)| e^{i\alpha(\theta)} P_{l'}(\cos \theta) \sin \theta d\theta &= \\ &= \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1)(e^{2i\delta_l}-1) P_l P_{l'} \sin \theta d\theta. \end{aligned}$$

Making use of the properties of orthogonality of the Legendre polynomials, we obtain

$$\int |f(\theta)| e^{i\alpha(\theta)} P_{l'}(\cos \theta) \sin \theta d\theta = \frac{1}{ik} (e^{2i\delta_{l'}} - 1). \quad (95.29)$$

Equating the real parts of relation (95.29), we find

$$|f(\theta)| \cos \alpha(\theta) P_l(\cos \theta) d \cos \theta = \frac{\sin 2\delta_l}{k}. \quad (95.30)$$

From formula (95.30) it is clear that the replacement of α by $\pi - \alpha$ leads to a change of sign of the left-hand side. To conserve the equality it is necessary to reverse the sign of all the phase shifts δ_l . Thus the uncertainty in the quantity α leads to an uncertainty in the sign of all the phase shifts.

If the sign of only one of the phase shifts is determined in an independent way, then the relation of δ_l to all the other phase shifts becomes unambiguous. The sign of one phase shift (namely, the s-wave one) can be established, for example, from the study of the scattering and interference of slow particles. It should be pointed out that although the calculations show us the possibility of determining the scattering amplitude, the solution of the integral equation (95.28) is a difficult problem.

§96. S-matrix and perturbation theory

If the total Hamiltonian can be written in the form of a sum $\hat{H} = \hat{H}_0 + \hat{H}'$ where \hat{H}_0 describes the behaviour of non-interacting particles and \hat{H}' their interaction, then to find the explicit form of the S-matrix it is convenient to make use of the interaction representation. The wave function in this representation is defined by eq. (49.21). The operator $\hat{V}(t, t_0)$ defined by formula (49.1), which transforms the wave function at a given instant of time t_0 into the wave function at the instant of time t , can also be defined in the interaction representation. That is, writing

$$\varphi(t) = \hat{V}(t, t_0) \varphi(t_0) \quad (96.1)$$

and substituting into (49.21), we find

$$i\hbar \frac{\partial \hat{V}(t, t_0)}{\partial t} = \hat{H}'_{\text{int}}(t) \hat{V}(t, t_0), \quad (96.2)$$

$$\hat{V}(t_0, t_0) = 1. \quad (96.3)$$

System (96.2) and (96.3) can be compared with the integral equation

$$\hat{V}(t, t_0) = 1 - \frac{i}{\hbar} \int_{t_0}^t dt' \hat{H}'_{\text{int}}(t') \hat{V}(t', t_0). \quad (96.4)$$

Integral equation (96.4) can be solved by a method of successive approximation

$$\begin{aligned} V(t, -\infty) = & 1 - \frac{i}{\hbar} \int_{-\infty}^t dt_1 \hat{H}'_{\text{int}}(t_1) + \\ & + \left(-\frac{i}{\hbar}\right)^2 \int_{-\infty}^t dt_1 \int_{-\infty}^{t_1} dt_2 \hat{H}'_{\text{int}}(t_1) \hat{H}'_{\text{int}}(t_2) + \dots \end{aligned} \quad (96.5)$$

The general term of the series is of the form

$$\hat{V}^{(n)} = \left(-\frac{i}{\hbar}\right)^n \int_{-\infty}^t dt_1 \int_{-\infty}^{t_1} dt_2 \dots \int_{-\infty}^{t_{n-1}} \hat{H}'_{\text{int}}(t_1) \hat{H}'_{\text{int}}(t_2) \dots \hat{H}'_{\text{int}}(t_n) dt_n. \quad (96.6)$$

It is evident that the range of integration over the variables t_1, t_2, \dots, t_n has the order

$$t_1 > t_2 > \dots > t_n. \quad (96.7)$$

In order to simplify the notation and so that one need not follow the order of carrying out the integration, it is convenient to symmetrize formula (96.6). In the case of a function symmetric with respect to its variables use can be made of the formula

$$\begin{aligned} \int_a^b dt_1 \int_a^{t_1} dt_2 \dots \int_a^{t_{n-1}} dt_n f(t_1, \dots, t_n) = \\ = \frac{1}{n!} \int_a^b dt_1 \int_a^b dt_2 \dots \int_a^b dt_n f(t_1, \dots, t_n). \end{aligned} \quad (96.8)$$

For the purpose mentioned we introduce the so-called chronological operator \hat{P} , which by definition arranges time-dependent operators in chronological sequence, i.e. in order of decreasing times (96.7):

$$\hat{P}\hat{L}(t_1)\hat{M}(t_2) = \begin{cases} \hat{L}(t_1)\hat{M}(t_2) & \text{for } t_1 > t_2 \\ \hat{M}(t_2)\hat{L}(t_1) & \text{for } t_2 > t_1. \end{cases} \quad (96.9)$$

A representation of this operator could, for example, be the expression

$$\hat{P} = \frac{1 + \epsilon(t_1 - t_2)}{2} + \frac{1 - \epsilon(t_2 - t_1)}{2},$$

where $\epsilon(x)$ is the so-called sign function

$$\epsilon(x) = \frac{|x|}{x} = \begin{cases} 1 & \text{for } x > 0, \\ -1 & \text{for } x < 0. \end{cases}$$

By means of the chronological operator we can write

$$\begin{aligned} \int_{-\infty}^t dt_1 \int_{-\infty}^{t_1} dt_2 \dots \int_{-\infty}^{t_{n-1}} dt_n \hat{H}'_{\text{int}}(t_1) \dots \hat{H}'_{\text{int}}(t_n) = \\ = \frac{1}{n!} \int_{-\infty}^t dt_1 \int_{-\infty}^{t_1} dt_2 \dots \int_{-\infty}^t dt_n \hat{P} \{ \hat{H}'_{\text{int}}(t_1) \dots \hat{H}'_{\text{int}}(t_n) \}. \end{aligned} \quad (96.10)$$

Hence for $\hat{V}(t, -\infty)$ we find

$$\begin{aligned} \hat{V}(t, -\infty) = 1 + \sum_{n=1} \left(-\frac{i}{\hbar} \right)^n \frac{1}{n!} \int_{-\infty}^t dt_1 \dots \int_{-\infty}^t dt_n \hat{P} \{ \hat{H}'_{\text{int}}(t_1) \dots \hat{H}'_{\text{int}}(t_n) \} = \\ = P \exp \left(-\frac{i}{\hbar} \int_{-\infty}^t \hat{H}'_{\text{int}}(t) dt \right). \end{aligned}$$

In accordance with the definition of the S -matrix

$$\begin{aligned} S = \lim_{\substack{t \rightarrow \infty \\ t_0 \rightarrow -\infty}} \hat{V}(t, t_0) = \lim_{t \rightarrow \infty} \hat{V}(t, -\infty) = \\ = \hat{P} \exp \left(-\frac{i}{\hbar} \int_{-\infty}^{\infty} \hat{H}'_{\text{int}}(t) dt \right). \end{aligned} \quad (96.11)$$

The formula obtained, called Dyson's formula, allows one to relate the S -matrix to the interaction energy \hat{H}' (if this last exists). It is accurate in the sense that the summation of the entire perturbation series is carried out in it.

It is easily seen that the first terms of the expansion of the general formula for the S -matrix lead to ordinary perturbation theory.

For ease of calculation we restrict ourselves to first order perturbation theory, writing

$$S^{(1)} = 1 - \frac{i}{\hbar} \int_{-\infty}^{\infty} dt \hat{H}'_{\text{int}}(t). \quad (96.12)$$

The operator \hat{P} is in this case identically equal to one. Taking the matrix element with respect to states $a \neq b$, which are eigenstates of the Hamiltonian H_0 , we have

$$S_{ba} = -\frac{i}{\hbar} \int_{-\infty}^{\infty} dt (\hat{H}'_{\text{int}})_{ba}.$$

Passing to the Schrödinger representation and making use of definition (49.19), we obtain

$$\begin{aligned} S_{ba}^{(1)} &= -\frac{i}{\hbar} \int_{-\infty}^{\infty} dt \langle b | e^{(i/\hbar)\hat{H}_0 t} \hat{H}' e^{-(i/\hbar)\hat{H}_0 t} | a \rangle = \\ &= -\frac{i}{\hbar} \langle b | \hat{H}' | a \rangle \int_{-\infty}^{\infty} \exp \left[\frac{i}{\hbar} (E_b - E_a) t \right] dt = -2\pi i \hat{H}'_{ba} \delta(E_b - E_a). \end{aligned}$$

We see that $S_{ba}^{(1)}$ is the same as the transition amplitude in the first approximation of perturbation theory.

Analogous, although more cumbersome calculations allow $S_{ba}^{(2)}$ to be identified with the transition amplitude in the second order of perturbation theory.

In spite of the convenience of the notation of Dyson's formula, which is often used in intermediate calculations, for the actual calculation of the S -matrix one has to carry out an expansion in a series and integration by terms.

An important feature of Dyson's formula is the fact that it can easily be transformed into a relativistically invariant form. Hence it is of particular importance in calculating relativistic effects.

§97. Analytic properties of the S -matrix

As we have already stressed, a number of important results of scattering theory which are not associated with the use of a particular form of the interaction potential can be obtained by means of the S -matrix technique. This is associated, in particular, with the analytic properties of the S -matrix*. For simplicity, we shall in what follows restrict ourselves to the case of elastic scattering. Then the elements of the S -matrix are given by formula (95.24).

As was shown in §35, the asymptotic expression for the radial component of the wave function of a particle of energy $E = \hbar^2 k^2 / 2m$ and angular momentum l , regular at the origin, has the form

$$\chi_{kl} = r R_{kl} = a_l(k) e^{i(kr - \frac{1}{2}l\pi)} + b_l(k) e^{-i(kr - \frac{1}{2}l\pi)}. \quad (97.1)$$

In deriving (97.1) it was assumed that the potential decreases more rapidly than r^{-1} at large distances.

Comparing expression (97.1) with (86.2) and taking into account (95.24), the matrix elements $S_{aa}^l \equiv S_l$ can be expressed in terms of the constants $a_l(k)$ and $b_l(k)$

$$S_l(k) = -\frac{a_l(k)}{b_l(k)}. \quad (97.2)$$

We shall now formally consider the wave function χ_{kl} and, correspondingly, the function $S_l(k)$ to be functions of the complex variable k . We shall show, first of all, that the function of a complex variable $S_l(k)$ should be given only in one quadrant and not in the entire plane of the complex variable k . Indeed, since the Schrödinger equation does not change under the replacement of k by $-k$, the function χ_{-kl} , by virtue of the uniqueness of the solution, describes the same state as the function χ_{kl} . These two functions can differ only by a constant factor. Replacing k by $-k$ in (97.1), we obtain

$$\frac{a_l(k)}{b_l(k)} = \frac{b_l(-k)}{a_l(-k)}.$$

Hence it follows that

$$S_l(k) = S_l^{-1}(-k). \quad (97.3)$$

* For a more detailed consideration of the problems touched upon in this section and for a bibliography see A.I.Baz, L.B.Zeldovich and A.M.Perelomov, *Rasseyanie, reaktsii i raspady v nerelyativistskoi kvantovoi mekhanike* (Scattering, reactions and decays in non-relativistic quantum mechanics) (Nauka, Moscow, 1966).

Further, we note that, since the Schrödinger equation is real, the function χ_{kl}^* also must be the same to within a constant as the function χ_{kl} . Hence it is again easily found that

$$S_l(k) = (S_l^*(k))^{-1}. \quad (97.4)$$

Formula (97.4) is obtained for real k . Carrying out analytic continuation to the entire plane of the complex variables k , we have

$$S_l(k) = (S_l^*(k^*))^{-1}. \quad (97.5)$$

Relations (97.3) and (97.5) connect the values of the function $S_l(k)$ given in one of the quadrants of the plane of the complex variable k with its values at the corresponding points of the remaining three quadrants. From relation (97.4) it follows that for real k , $|S_l(k)|^2 = 1$, i.e. the phase shift δ_l is real ($S_l = e^{2i\delta_l}$). On the contrary, as is seen from (97.5) and (97.3), the function $S_l(k)$ is real on the imaginary axis, so that the phase shift δ_l is imaginary.

Let us consider the position of the singularities of the function $S_l(k)$. We assume that there corresponds to the potential $U(r)$ a bound state of the particle with energy $-E_0$. The bound state is described by the wave function $\chi_{k_0 l}$ regular at the origin and falling off at large distances as $e^{-|k_0| r}$, where $k_0 = i(2m\hbar^{-2}|E_0|)^{\frac{1}{2}}$. Consequently, the function χ_{kl} analytically continued to the complex plane must fall off for $k = k_0$ as $e^{-|k_0| r}$. Hence the relation $b_l(k_0) = 0$ must be satisfied at the point $k = k_0$. In accordance with formula (97.2), the function $S_l(k)$ has a pole at the point $k = k_0$. As follows from (97.3), the function $S_l(k)$ reduces to zero at the symmetric point lying in the lower half-plane, i.e. at $k = -k_0$. Thus we arrive at the conclusion that to each bound state there corresponds a pole of the function $S_l(k)$ lying at the corresponding point of the upper imaginary semiaxis in the plane of the complex variable k . It should be noted that so-called 'false' poles, which do not correspond to any stationary state, may also arise on the imaginary semiaxis. It can be shown (see ref. on p. 412) that 'false' poles do not arise when the so-called cut-off radius R is introduced, i.e. when the condition $U(r) = 0$ is introduced for $r > R$, where the radius R may be as large as one wants.

We note that the function $S_l(k)$ cannot have poles in the upper half-plane lying anywhere off the imaginary axis. Indeed, to such a pole there would correspond a complex value of the energy of the bound state, which is impossible.

The function $S_l(k)$ may also have poles in the lower half-plane, and there they may also lie off the imaginary semiaxis. As follows immediately from relations (97.3) and (97.5), these poles must be situated in pairs symmetric with respect to the imaginary semiaxis. In the upper half-plane there corre-

spond to these poles zeros of the function $S_l(k)$. It is easily seen that to poles lying in the lower half-plane there correspond wave functions exponentially increasing at large distances. Such wave functions cannot, of course, correspond to a bound state. It can be shown that to poles in the lower half-plane there correspond quasi-stationary states of the system, i.e. states which decay in the course of a certain finite time T .

We find the residue of the function $S_l(k)$ with respect to the pole to which there corresponds a bound state with energy $E = -E_0$ or the value $k = k_0 = i(2m\hbar^{-2}|E_0|)^{\frac{1}{2}}$. Denoting this residue by c_l , we write the function S_l in the neighbourhood of the point $k = k_0$ in the form

$$S_l = \frac{c_l}{k - k_0}. \quad (97.6)$$

The quantity c_l is connected by a simple relation with the amplitude of the wave function corresponding to the stationary state with energy $E = -E_0$. In order to establish this relation we write down equations satisfied by the function χ_{kl} and by its derivative with respect to energy

$$\begin{aligned} \chi_{kl}'' + \frac{2m}{\hbar^2} \left(E - U - \frac{\hbar l(l+1)}{2mr^2} \right) \chi_{kl} &= 0 \\ \left(\frac{\partial \chi_{kl}}{\partial E} \right)'' + \frac{2m}{\hbar^2} \left(E - U - \frac{\hbar^2 l(l+1)}{2mr^2} \right) \frac{\partial \chi_{kl}}{\partial E} &= -\frac{2m}{\hbar^2} \chi_{kl}. \end{aligned}$$

We shall assume the function χ_{kl} to be normalized by the condition

$$\int_0^\infty |\chi_{kl}|^2 dr = 1.$$

Multiplying the first equation by $\partial \chi_{kl} / \partial E$ and the second by χ_{kl} , subtracting one from the other and integrating with respect to dr , we obtain

$$\chi_{kl}' \frac{\partial \chi_{kl}}{\partial E} - \chi_{kl} \left(\frac{\partial \chi_{kl}}{\partial E} \right)' = \frac{2m}{\hbar^2} \int_0^r \chi_{kl}^2 dr. \quad (97.7)$$

We apply this relation for $E = -E_0$ and $r \rightarrow \infty$. Expanding the functions $a_l(k)$ and $b_l(k)$ near the point $k = k_0$ in a series and rename the constant

$$a_l(k) = a_l(k_0) = A_l i^{-l}, \quad b_l(k) = \beta_l(k - k_0). \quad (97.8)$$

Making use of these expansions and of relations (97.7) and (97.1), we obtain

$$\beta_l = -\frac{i}{a_l} = -\frac{i^{l+1}}{A_l}. \quad (97.9)$$

Substituting expression (97.9) into (97.2), the residue c_l at the point $k = k_0$ is

$$c_l = iA_l^2(-1)^{l+1}. \quad (97.10)$$

Thus we have related the value of the residue c_l to the amplitude A_l in the asymptotic expression of the wave function $\chi_{kl} = A_l e^{-ik_0 r}$ of the bound state.

The study of the behaviour of the scattering phase shifts $\delta_l(k)$ and, consequently, also of the function $S_l(k) = e^{2i\delta_l(k)}$, and the extrapolation of these results to the complex region make it possible, on the basis of (97.10), also to draw definite conclusions concerning the wave function of the bound state.

The analytic properties of the quantities $S_l(k)$ make it possible to obtain important relations which must be satisfied by the scattering amplitude. These relations are called dispersion relations. Dispersion relations for the amplitude of the scattering at zero angle $f(0, k)$ are the simplest and at the same time the most important. Dispersion relations establish the connection between the real and imaginary parts of the scattering amplitude $f(\vartheta, k)$

$$f(\vartheta, k) = \operatorname{Re} f(\vartheta, k) + i \operatorname{Im} f(\vartheta, k)$$

and are based on the use of the Cauchy formula in the theory of analytic functions.

Suppose that $F(k)$ is a certain function which is analytic in the upper half-plane of the complex variable k and has simple poles on the upper imaginary axis. Let us consider the integral

$$\int_C \frac{F(k') dk'}{k' - k}$$

taken over the contour shown in fig. V.28.

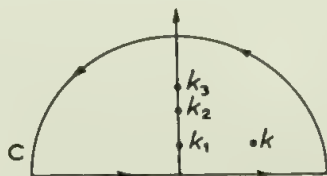


Fig. V.28

The integral is determined by the sum of the residues of the integrand. These residues are taken at the point $k' = k$ and at the points $k' = k_1, k_2, \dots$, where the poles of the function $F(k)$ lie. If the function $F(k)$ tends to zero sufficiently rapidly as $|k| \rightarrow \infty$, then the integral over the upper semicircle is equal to zero. We then have

$$\int_{-\infty}^{\infty} \frac{F(k') dk'}{k' - k} = 2\pi i \left(F(k) + \sum_n \frac{\text{Res } F(k_n)}{k_n - k} \right). \quad (97.11)$$

Here $\text{Res } F(k_n)$ denotes the residue of the function F at the point $k' = k_n$. Now let the imaginary part of k tend to zero, so that k tends to the point k_0 lying on the real axis. In this case

$$\int_{-\infty}^{\infty} \frac{F(k') dk'}{k' - k_0} = P \int_{-\infty}^{\infty} \frac{F(k') dk'}{k' - k_0} + i\pi F(k_0). \quad (97.12)$$

Here P denotes that the integral is understood in the sense of the principal value

$$P \int_{-\infty}^{\infty} \frac{F(k') dk'}{k' - k_0} = \lim_{\epsilon \rightarrow 0} \left[\int_{-\infty}^{k_0 - \epsilon} \frac{F(k') dk'}{k' - k_0} + \int_{k_0 + \epsilon}^{\infty} \frac{F(k') dk'}{k' - k_0} \right],$$

and the second term on the right-hand side of (97.12) arises from the integration over a small semicircle around the point $k = k_0$.

Based on the results (97.11) and (97.12) we obtain the dispersion relation for the amplitude of the scattering at zero angle $f(0, k)$. This amplitude is connected with the matrix elements S_l (95.24) by relation (86.11)

$$f(0, k) = \frac{1}{2ik} \sum_l (2l+1)(S_l - 1). \quad (97.13)$$

From this expression it follows that the poles of the function $S_l(k)$ are also the poles of the function $f(0, k)$, and the function $f(0, k)$ has no other poles. The point $k = 0$ is not a pole at all, since for $k \rightarrow 0$, $\delta \rightarrow 0$, $S_l \rightarrow 1$ (see §86). Thus the function $f(0, k)$ is analytic in the upper half-plane of the complex variable k and has poles on the upper imaginary semiaxis. Dispersion relations for this function are easily obtained if one substitutes into relations (97.11) and (97.12) the function $F(k)$ in the form

$$F(k) = f(0, k) - f(0, \infty). \quad (97.14)$$

From the amplitude $f(0, k)$ one subtracts its value for $k \rightarrow \infty$ in order to reduce the integral over the large semicircle to zero (fig. V.28). For $k \rightarrow \infty$ the term with the potential $U(r)$ can be neglected in the Schrödinger equation. The solution of such an equation has the form of a plane wave. Substituting such a solution into (83.10), we obtain

$$f(0, \infty) = -\frac{m}{2\pi\hbar^2} \int U(r) dV. \quad (97.15)$$

Expression (97.15) represents the scattering amplitude in the Born approximation (see §84), i.e. $f(0, \infty) = f_B$. Substituting (97.14) into relations (97.11) and (97.12) and taking into account that the integral with the Born amplitude reduces to zero, we obtain

$$f(0, k) = f_B + \frac{1}{i\pi} P \int_{-\infty}^{\infty} \frac{f(0, k') dk'}{k' - k} - 2 \sum_{n,l} \frac{\text{Res } f(0, k_{nl})}{k_{nl} - k}. \quad (97.16)$$

In this relation k is assumed to be real, and the index zero is dropped. The points $k = k_{nl}$ lie on the upper imaginary semiaxis and correspond to the poles of the function $S_l(k)$. The summation in (97.16) is carried out over all bound states. The expression for the residue of the function S_l in terms of the amplitude of the corresponding bound state is given by formula (97.10). Taking into account (97.13), we have

$$\text{Res } f(0, k_{nl}) = \frac{1}{2k_{nl}} A_{nl}^2 (-1)^{l+1} (2l+1). \quad (97.17)$$

Relation (97.16) can be rewritten in a somewhat different form, if it is taken into account, according to (97.3) and (97.4), that for real k $S_l(-k) = S_l^*(k)$ and, correspondingly (see (97.13)), that $f(0, -k) = f^*(0, k)$. Hence the integration in (97.16) can be carried out only over positive values of k , having physical meaning. Equating the real parts on the left and the right in (97.16), we have finally

$$\text{Re } f(0, k) = f_B + \frac{2}{\pi} P \int_0^{\infty} \frac{\text{Im } f(0, k') k' dk'}{k'^2 - k^2} - \text{Re} \sum_{n,l} \frac{A_{nl}^2 (-1)^{l+1} (2l+1)}{k_{nl}(k_{nl} - k)}. \quad (97.18)$$

The imaginary part of the amplitude $\text{Im } f(0, k)$ involved in the right-hand side of the equation can be expressed, according to the optical theorem (see (91.11)), in terms of a physically observed quantity; the total scattering cross section $\sigma(k)$. Hence also the real part, $\text{Re } f(0, k)$, according to (97.18) can be expressed in terms of physically observable quantities. Dispersion relations are

at present widely used. In particular, by means of them one can immediately remove the ambiguity (noted in §95) in the choice of phase shifts for a known law of scattering, i.e. for a known cross section. We stress that dispersion relations are based on as general a property of the S -matrix as its analyticity which results from the causality principle.

§98. Time reversal and the principle of detailed balance

Let us consider the properties of the S -matrix associated with the symmetry of the Schrödinger equation with respect to time reversal. We have already touched upon this question in §6 and shall now consider it in more detail.

The symmetry with respect to time reversal means that there exists a solution $\psi_{\text{rev}}(x, t)$ of the 'reversed' Schrödinger equation expressed in terms of the function $\psi(x, -t)$. If the operator \hat{H} does not depend explicitly on time, then

$$i\hbar \frac{\partial \psi^*(x, -t)}{\partial t} = \hat{H}^* \psi^*(x, -t). \quad (98.1)$$

For $\hat{H}^* = \hat{H}$ eq. (98.1) is the same as the initial equation (27.7), and the function $\psi^*(x, -t)$ describes the process reversed in time (see (6.9)). In the more general case (a charged particle in a magnetic field) we have to set

$$\psi_{\text{rev}}(x, t) = \hat{V} \psi^*(x, -t), \quad (98.2)$$

where \hat{V} is a certain operator. Operating on eq. (98.1) from the left with the operator \hat{V} , we obtain the equation for the function ψ_{rev}

$$i\hbar \frac{\partial \psi_{\text{rev}}(x, t)}{\partial t} = \hat{V} \hat{H}^* \hat{V}^{-1} \psi_{\text{rev}}(x, t). \quad (98.3)$$

This equation is the same as the initial Schrödinger equation (27.7) under the condition

$$\hat{V} \hat{H}^* = \hat{H} \hat{V}. \quad (98.4)$$

From the Hermitian property of the operator \hat{H} it follows that the operator \hat{V} must be unitary, i.e. $\hat{V}^{-1} = \hat{V}^\dagger$. To the law of transformation of wave functions (98.2) there corresponds a definite law of transformation of arbitrary operators \hat{F} . This law can be found by the usual methods (see §46, 48.49).

A lack of generality arises in the case given only in connection with the

fact that the operator \hat{V} operates not on the function ψ but on the function ψ^* . We shall find the operator \hat{F}_{rev} (reversed in time) proceeding from the requirement that the matrix element of the operator \hat{F} taken with respect to the functions ψ_{rev} must be the same as the matrix elements of the operator \hat{F}_{rev} taken with respect to the functions $\psi(x, -t)$

$$\langle \psi_{\text{rev}} | \hat{F} | \psi_{\text{rev}} \rangle = \langle \psi(-t) | \hat{F}_{\text{rev}} | \psi(-t) \rangle. \quad (98.5)$$

Making use of relation (98.2), we obtain

$$\langle \psi_{\text{rev}} | \hat{F} | \psi_{\text{rev}} \rangle = \langle \hat{V} \psi^*(-t) | \hat{F} | \hat{V} \psi^*(-t) \rangle = \langle \psi^*(-t) | \hat{V}^\dagger \hat{F} \hat{V} | \psi^*(-t) \rangle.$$

Hence it follows (see (17.3)) that

$$\tilde{\hat{F}}_{\text{rev}} = \hat{V}^\dagger \hat{F} \hat{V}, \quad (98.6)$$

where $\tilde{\hat{F}}_{\text{rev}}$ denotes the transpose of the operator \hat{F}_{rev} . As is easily seen from (98.4) and (98.6), the operator \hat{H} is invariant under time reversal, i.e. $\hat{H}_{\text{rev}} = \hat{H}$. Here we have made use of the condition of hermiticity of the Hamiltonian $\hat{H} = \hat{H}^*$. Relation (98.6) can serve as a basis for finding the operator \hat{V} . Indeed, it is natural to require that the quantum operators transform under time reversal in the same way as the corresponding classical quantities. Quantities such as energy, coordinates, electric field strength and so on are invariant under time reversal. The corresponding operators also must be invariant. Velocity, momentum, angular momentum, magnetic field strength and so on change sign under time reversal. The corresponding operators must have the same property. For example, the relations

$$\hat{\mathbf{r}}_{\text{rev}} = \hat{\mathbf{r}}, \quad \hat{\mathbf{p}}_{\text{rev}} = -\hat{\mathbf{p}}, \quad \hat{\mathbf{L}}_{\text{rev}} = -\hat{\mathbf{L}} \quad (98.7)$$

must be fulfilled. The spin transforms as the angular momentum, i.e. the following relation must be fulfilled:

$$\hat{\mathbf{S}}_{\text{rev}} = -\hat{\mathbf{S}}. \quad (98.8)$$

Let us consider, for example, a particle with spin $\frac{1}{2}$. Proceeding from relations (98.8) it is easy to find the operator \hat{V}_s operating on the spin variables under time reversal. Making use of expression (98.6) and taking into account the form of the spin operators (60.15) and (60.16), we have

$$\hat{V}_s^\dagger \hat{s}_x \hat{V}_s = -\hat{s}_x, \quad \hat{V}_s^\dagger \hat{s}_y \hat{V}_s = \hat{s}_y, \quad \hat{V}_s^\dagger \hat{s}_z \hat{V}_s = -\hat{s}_z. \quad (98.9)$$

From these relations by means of (60.12) we easily find

$$\hat{V}_s = i\sigma_y. \quad (98.10)$$

(We have chosen the phase factor in such a way that the operator \hat{V}_s is real.)

For the motion of a particle in a magnetic field the operator \hat{V} must involve changing the direction of the magnetic field (or of the vector potential \mathbf{A}) to the opposite direction. Taking this fact into account, relation (98.4) has the form

$$\sigma_y \hat{H}^*(-\mathbf{A}) = \hat{H}(\mathbf{A}) \sigma_y. \quad (98.11)$$

It is easily verified that the Hamiltonian \hat{H} (see (63.3)) satisfies this relation. The invariance of the Schrödinger equation under time reversal means that one can always find an operator \hat{V} satisfying condition (98.4) (for more details see the reference below*). However, the discovery in 1964 of the anomaly in the decay of K-mesons shows that under certain conditions the principle of time reversal may apparently be violated.

From the invariance of the Hamiltonian \hat{H} under the replacement $t \rightarrow -t$ there results the invariance of the S -matrix, i.e. (see (98.6)) the following relation holds:

$$\hat{V}^\dagger \hat{S} \hat{V} = \tilde{\hat{S}}. \quad (98.12)$$

The validity of this relation is easily checked, taking into account (98.4), for the operator $\hat{V}(t, t_0)$ (see (96.5)). Since the operator \hat{S} is defined as the limit of the operator $\hat{V}(t, t_0)$ (see (96.11)), it also satisfies relation (98.12).

Based on relation (98.12), it is easy to establish the relation directly between the matrix elements of the S -matrix for the direct and inverse reactions. We denote by ψ_a and ψ_b the wave functions of the initial and final states of the system. Then, taking into account (17.3), (98.2) and (98.12), we have

$$\langle \psi_b | \hat{S} | \psi_a \rangle = \langle \psi_a^* | \tilde{\hat{S}} | \psi_b^* \rangle = \langle \psi_a^* | \hat{V}^\dagger \hat{S} \hat{V} | \psi_b^* \rangle = \langle \hat{V} \psi_a^* | \hat{S} | \hat{V} \psi_b^* \rangle = \langle \psi_{a^*} | \hat{S} | \psi_{b^*} \rangle, \quad (98.13)$$

where ψ_{a^*} and ψ_{b^*} denote the 'reversed' wave functions of the states a and b . Thus the following equality is fulfilled:

$$S_{ba} = S_{a^*b^*}. \quad (98.14)$$

Relation (98.14) establishes the connection between the matrix elements of the S -matrix of the direct and 'reversed' processes. The states ψ_{a^*} and ψ_{b^*} differ from the states ψ_a and ψ_b by the sign of quantities such as velocities, momenta, angular momentum components, spin components and so on. Relation (98.14) or the equivalent relation (98.13) is called the reciprocity theorem. On the basis of this theorem a relation can be established between

* A.M.Baldin, V.I.Goldanskii and I.L.Rozenthal, *Kinematics of nuclear reactions* (Pergamon Press, Oxford, 1961).

the cross sections for direct and inverse reactions (principle of detailed balance).

Let us consider the reaction



We denote by $j_a, m_a, j_A, m_A, j_b, m_b, j_B, m_B$ respectively the total angular momenta and their components of the particles taking part in the reaction. According to (95.14), the cross sections for the direct and inverse reactions expressed in terms of the matrix elements of the S -matrix have the form

$$\frac{d\sigma_{ba}}{d\Omega_b} = \frac{4\pi^2}{k_a^2} |\langle j_b, m_b, j_B, m_B; -\mathbf{n}_b | \hat{S} | j_a, m_a, j_A, m_A; \mathbf{n}_a \rangle|^2, \quad (98.15)$$

$$\frac{d\sigma_{ab}}{d\Omega_a} = \frac{4\pi^2}{k_b^2} |\langle j_a, m_a, j_A, m_A; -\mathbf{n}_a | \hat{S} | j_b, m_b, j_B, m_B; \mathbf{n}_b \rangle|^2. \quad (98.16)$$

Since the momentum vector of the relative motion of the particles in the final state is directed away from the centre of mass, it is assigned a minus sign.

The relation between these cross sections cannot be written directly, since the reciprocity theorem relates the cross section for the direct process to that of the 'reversed' process which differs from (98.16) by the change of the signs of the angular momentum components m_a, m_A, m_b, m_B into the opposite signs. However, one can write the relation between averaged cross sections, i.e. cross sections summed over the components of the angular momenta of the final states and averaged over the components of the angular momenta of the initial states. Such cross sections no longer depend on the angular momentum components, and for them the reciprocity theorem (98.14) gives

$$\frac{1}{k_b^2} (2j_a+1)(2j_A+1) \frac{\overline{d\sigma_{ba}}}{d\Omega_b} = \frac{1}{k_a^2} (2j_b+1)(2j_B+1) \frac{\overline{d\sigma_{ab}}}{d\Omega_a}, \quad (98.17)$$

where

$$\frac{\overline{d\sigma_{ba}}}{d\Omega_b} = \frac{1}{(2j_a+1)(2j_A+1)} \sum_{m_a, m_A} \frac{d\sigma_{ba}}{d\Omega_b} \quad (98.18)$$

$$m_b, m_B$$

and

$$\overline{\frac{d\sigma_{ab}}{d\Omega_a}} = \frac{1}{(2j_b+1)(2j_B+1)} \sum_{\substack{m_a, m_A \\ m_b, m_B}} \frac{d\sigma_{ab}}{d\Omega_a}. \quad (98.19)$$

A relation analogous to (98.17) can also be written for the total cross sections

$$k_a^2(2j_a+1)(2j_A+1)\bar{\sigma}_{ba} = k_b^2(2j_b+1)(2j_B+1)\bar{\sigma}_{ab}. \quad (98.20)$$

We note also that the relation between non-averaged cross sections for the direct and inverse reactions can be established within the framework of applicability of perturbation theory:

$$\frac{1}{k_b^2} \frac{d\sigma_{ba}}{d\Omega_b} = \frac{1}{k_a^2} \frac{d\sigma_{ab}}{d\Omega_a}. \quad (98.21)$$

Indeed, in this case the transition probability and, consequently, also the cross section for the process, is determined by the square of the modulus of the matrix element of the perturbation Hamiltonian H'_{ba} , for which, by virtue of hermiticity, the relation $|H'_{ba}|^2 = |H'_{ab}|^2$ is fulfilled. From this equality there results relation (98.21).

The Method of Second Quantization and Radiation Theory

§99. Second quantization for systems of bosons and fermions*

One of the important formal mathematical methods often used in the quantum mechanics of a system of many particles is the so-called second-quantization method.

In this method a transition from the coordinate representation of the wave function to new variables is carried out. As new variables the numbers of particles in a given quantum state are chosen. Thus the system of particles is now characterized not by defining the wave function $\psi(\xi_1, \xi_2, \dots, \xi_N, t)$ but by defining a new function $c(n_1, n_2, \dots, t)$, where n_1, n_2, \dots are the numbers of particles in the 1st, 2nd and so on states. We shall call the quantities n_1, n_2, \dots the occupation numbers.

The quantity

$$|c(n_1, n_2, \dots, n_k, \dots, t)|^2 \quad (99.1)$$

gives the probability that at instant of time t there are n_1 particles in the first state, n_2 particles in the second state and so on. The second-quantization method turns out to be very convenient for those systems in which the

* In this section we follow L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

number of particles in a given state changes, and the production and disappearance of particles of a given kind occurs (for example, in the emission and absorption of photons, or in the β -decay of nuclei). The transition from the ordinary description to second quantization is an example of a transformation from one representation to another.

Let us formally consider a system of non-interacting identical particles. We shall first assume that the particles obey Bose statistics.

We denote by $\psi_1(\xi)$, $\psi_2(\xi)$, ..., $\psi_k(\xi)$ the whole set of orthogonal and normalized wave functions of an individual particle forming a complete system of functions chosen in an arbitrary way. The index k denotes the set of four quantum numbers characterizing the state of the particle. We pass to the representation in which the occupation numbers n_k and not the coordinates, ξ_i , of the particles are chosen as independent variables.

In the new representation the basis functions (see §65) are the symmetrized and normalized products of the wave functions $\psi_k(\xi_i)$ of the individual particles. Formula (65.5) for the general case where n_1 particles are in state ψ_1 , n_2 particles in state ψ_2 and so on assumes the form

$$\begin{aligned} \psi_{n_1, n_2, n_3, \dots}(\xi_1, \xi_2, \dots, \xi_N) = \\ = \left(\frac{n_1! n_2! n_3! \dots}{N!} \right)^{\frac{1}{2}} \sum \psi_{k_1}(\xi_1) \psi_{k_2}(\xi_2) \dots \psi_{k_N}(\xi_N). \end{aligned} \quad (99.2)$$

The summation is carried out only over all permutations of different indices k_1, k_2, \dots .

We introduce the operators \hat{a}_k^\dagger and \hat{a}_k which act on the new variables, the occupation numbers in state k . We define these operators by the formulae

$$\hat{a}_k \psi_{n_1, \dots, n_k, \dots} = (n_k)^{\frac{1}{2}} \psi_{n_1, \dots, n_k-1, \dots}, \quad (99.3)$$

$$\hat{a}_k^\dagger \psi_{n_1, \dots, n_k, \dots} = (n_k+1)^{\frac{1}{2}} \psi_{n_1, \dots, n_k+1, \dots}. \quad (99.4)$$

The operator \hat{a}_k reduces the number of particles in state k by one, i.e. it replaces n_k by n_k-1 . The operator \hat{a}_k^\dagger increases this number by one, i.e. it replaces n_k-1 by n_k . It is obvious that the consecutive application of the operators \hat{a}_k and \hat{a}_k^\dagger does not change the number of particles in the k th state, i.e.

$$\hat{a}_k^\dagger \hat{a}_k \psi_{n_1, \dots, n_k, \dots} = n_k \psi_{n_1, \dots, n_k, \dots}. \quad (99.5)$$

The matrix elements of the operators \hat{a}_k and \hat{a}_k^\dagger are of the form

$$\langle n_1, n_2, \dots, n_k-1, \dots | \hat{a}_k | n_1, n_2, \dots, n_k, \dots \rangle = (a_k)_{n_k-1, n_k} = n_k^{\frac{1}{2}}, \quad (99.6)$$

$$\langle n_1, n_2, \dots, n_k+1, \dots | \hat{a}_k^\dagger | n_1, n_2, \dots, n_k, \dots \rangle = (a_k^\dagger)_{n_k+1, n_k} = (n_k+1)^{\frac{1}{2}}, \quad (99.7)$$

$$(a_k^\dagger a_k)_{n_k, n_k} = n_k \delta_{n_k, n_k}. \quad (99.8)$$

In accordance with their meaning, the operators \hat{a}_k and \hat{a}_k^\dagger are called respectively the annihilation and creation operators of a particle in the k th state. The operator $\hat{a}_k^\dagger \hat{a}_k$ is called the operator of the number of particles n_k in the state k .

We have already encountered operators similar to the operators \hat{a}_k and \hat{a}_k^\dagger in § 50 in considering the problem of the harmonic oscillator. It is easily seen that the operators \hat{a}_k and \hat{a}_k^\dagger satisfy the commutation relations

$$\begin{aligned} \hat{a}_k \hat{a}_l^\dagger - \hat{a}_l^\dagger \hat{a}_k &= \delta_{kl}, \\ \hat{a}_k \hat{a}_l - \hat{a}_l \hat{a}_k &= 0, \\ \hat{a}_k^\dagger \hat{a}_l^\dagger - \hat{a}_l^\dagger \hat{a}_k^\dagger &= 0. \end{aligned} \quad (99.9)$$

We shall show how the ordinary operators acting on a wave function in the coordinate representation can be expressed in terms of the creation and annihilation operators of particles, i.e. in the second-quantization representation.

Let us consider the operator $\hat{L}(\xi_i)$ acting on the coordinates of the i th particle. The coordinates are understood to include the spin coordinates. Since all the particles are equivalent, we introduce the operator $\hat{L}_1 = \sum_{i=1}^N \hat{L}(\xi_i)$. Let us find the expression for it in the second-quantization representation. We obtain the matrix elements of \hat{L}_1 by means of the basis functions (99.2).

We have by definition

$$\begin{aligned} \langle n'_1, \dots, n'_k, \dots | \hat{L}_1 | n_1, \dots, n_k, \dots \rangle &= \\ &= \left\langle n'_1, \dots, n'_k, \dots \left| \sum_{i=1}^N \hat{L}(\xi_i) \right| n_1, \dots, n_k, \dots \right\rangle. \end{aligned} \quad (99.10)$$

Let us consider one term of the sum over the particles

$$\begin{aligned} \langle n'_1, \dots, n'_k, \dots | \hat{L}(\xi_i) | n_1, \dots, n_k, \dots \rangle &= \\ &= \int \psi_{n'_1, \dots, n'_k, \dots}^* \hat{L}(\xi_i) \psi_{n_1, \dots, n_k, \dots} d\xi_1 \dots d\xi_N. \end{aligned} \quad (99.11)$$

(Summation over spin variables is implied.) The operator $\hat{L}(\xi_i)$ acts only on the variables of the i th particle. Hence we can write

$$\hat{L}(\xi_i)\psi_{n_1, \dots, n_k, \dots} = \left(\frac{n_1! \dots n_k! \dots}{N!} \right)^{\frac{1}{2}} \sum \psi_{k_1}(\xi_1) \dots \psi_{k_N}(\xi_N) \hat{L}(\xi_i) \psi_{k_i}(\xi_i). \quad (99.12)$$

Multiplying (99.12) by the function $\psi_{n_1, \dots}^*$ and integrating we note, first of all, that the integrals over all variables except ξ_i contain only the products of wave functions.

By virtue of orthogonality of the latter all integrals involving factors of the form $\psi_1^*(\xi_1)\psi_2(\xi_1)$, i.e. containing the products of the wave functions of the particles (except of the i th particle) referring to different states, will reduce to zero.

In the double sum over permutations (99.11) only those terms which contain the products of the wave functions of the particles (except the i th particle) referring to the same states differ from zero. The integral over the variables ξ_i is of the form

$$(\hat{L}(\xi_i))_{lk} = \int \psi_l^*(\xi_i) \hat{L}(\xi_i) \psi_k(\xi_i) d\xi_i.$$

This means that for $l \neq k$ a transition of the particle takes place from the k th state into the l th state. Consequently, the number of particles in the k th state decreases by one, and in the l th state increases by one. We denote the corresponding matrix element by

$$\langle n_k-1, n_l | \hat{L}(\xi_i) | n_k, n_l-1 \rangle \quad (99.13)$$

(the operator is diagonal with respect to other occupation numbers and we do not write them down). The functions involved in the matrix element are of the form

$$\begin{aligned} \psi_{n_1, \dots, n_k-1, \dots}^* &= \left(\frac{n_1! \dots (n_k-1)! \dots n_l! \dots}{N!} \right)^{\frac{1}{2}} \sum \psi_{k_1}^*(\xi_1) \dots \psi_{k_N}^*(\xi_N), \\ \psi_{n_1, \dots, n_l-1, \dots} &= \left(\frac{n_1! \dots n_k! \dots (n_l-1)! \dots}{N!} \right)^{\frac{1}{2}} \sum \psi_{k_1}(\xi_1) \dots \psi_{k_N}(\xi_N). \end{aligned}$$

By virtue of the orthogonality of the wave functions integration over the coordinates of all particles gives (taking into account the permutations of $N-1$ particles excluding the i th particle)

$$\begin{aligned}
\langle n_k-1, n_l | \hat{L}(\xi_i) | n_k, n_l-1 \rangle &= \left(\frac{n_1! \dots (n_k-1)! \dots n_l! \dots}{N!} \right)^{\frac{1}{2}} \left(\frac{n_1! \dots n_k! \dots (n_l-1)! \dots}{N!} \right)^{\frac{1}{2}} \times \\
&\times \frac{(N-1)!}{n_1! \dots (n_k-1)! \dots (n_l-1)! \dots} (\hat{L}(\xi_i))_{lk} \\
&= \frac{(n_k n_l)^{\frac{1}{2}}}{N} (\hat{L}(\xi_i))_{lk}.
\end{aligned}$$

Since the operators $\hat{L}(\xi_1), \hat{L}(\xi_2), \dots$ differ from each other only in the number of particles on whose coordinates they act, all matrix elements differing in the number of particles are equal to each other. Hence for the matrix element (99.10) of the operator \hat{L} we can finally write

$$\begin{aligned}
\langle n_k-1, n_l | \hat{L}_1 | n_k, n_l-1 \rangle &= \left\langle n_k-1, n_l \left| \sum_{i=1}^N \hat{L}(\xi_i) \right| n_k, n_l-1 \right\rangle = \\
&= N \langle n_k-1, n_l | \hat{L}(\xi_i) | n_k, n_l-1 \rangle = (n_k n_l)^{\frac{1}{2}} (L(\xi))_{lk}. \quad (99.14)
\end{aligned}$$

In the case where the diagonal matrix element is considered, i.e. where the distribution of the number of particles over states does not change, we have analogously

$$\langle n_1, n_2, \dots | \hat{L}_1 | n_1, n_2, \dots \rangle = \sum_k n_k (L(\xi))_{kk}. \quad (99.15)$$

We now introduce the operators \hat{a}^\dagger and \hat{a} into formulae (99.14) and (99.15). Then the operator \hat{L}_1 can be written in the form

$$\hat{L}_1 = \sum_{k,l} (L(\xi))_{lk} \hat{a}_l^\dagger \hat{a}_k. \quad (99.16)$$

Indeed, the matrix elements of this operator are, by virtue of (99.6) and (99.8), the same as the matrix elements (99.14) and (99.15).

An analogous result can be obtained in the same way for operators which act on the coordinates of two particles ξ_i and ξ_j .

The operator

$$\hat{L}_2 = \sum_{i,j \neq 1} \hat{L}(\xi_i, \xi_j)$$

is expressed in the second-quantization representation by the formula

$$\hat{L}_2 = \sum_{k,p,l,m} \langle l,m | \hat{L}(\xi, \xi') | k,p \rangle \hat{a}_m^\dagger \hat{a}_k^\dagger \hat{a}_p \hat{a}_l, \quad (99.17)$$

where the matrix elements are equal to

$$\langle l,m | \hat{L}(\xi, \xi') | k,p \rangle = \int \psi_l^*(\xi) \psi_m^*(\xi') \hat{L}(\xi, \xi') \psi_k(\xi) \psi_p(\xi') d\xi d\xi'. \quad (99.18)$$

By means of the general formulae (99.16) and (99.17) one can write the Hamiltonian of a system of particles in the second-quantization representation. In the case of a system of non-interacting particles in a given external field we have

$$\hat{H} = \sum_{i=1}^N \hat{H}_i = \sum_{i=1}^N (\hat{T}_i + U(\xi_i)) = \sum_{i=1}^N \left(-\frac{\hbar^2}{2m} \nabla_i^2 + U(\xi_i) \right), \quad (99.19)$$

where $U(\xi_i)$ is the potential energy of the i th particle in the external field, and \hat{T}_i is its kinetic energy operator. Operator (99.19) is evidently a particular case of the operator \hat{L}_1 . Correspondingly we can immediately write operator (99.19) in the second-quantization representation

$$\hat{H} = \sum_{k,l} (\hat{H}_i)_{lk} \hat{a}_l^\dagger \hat{a}_k. \quad (99.20)$$

Choosing as ψ_k the eigenfunctions of the Hamiltonian \hat{H}_i of an individual particle, we have

$$(H_i)_{lk} = \int \psi_l^*(\xi) \hat{H}_i(\xi) \psi_k(\xi) d\xi = E_k \delta_{lk},$$

where E_k is the energy of the particle in the k th state.

Hence, finally,

$$\hat{H} = \sum_k E_k \hat{a}_k^\dagger \hat{a}_k. \quad (99.21)$$

The energy of a system of particles is, by virtue of (99.8), equal to

$$E = \sum_k E_k n_k. \quad (99.22)$$

If the eigenfunctions of the operator \hat{T}_i corresponding to the eigenvalues ϵ_k are chosen as ψ_k , then (99.20) is rewritten in the form

$$\hat{H} = \sum_k \epsilon_k \hat{a}_k^\dagger \hat{a}_k + \sum_{k,l} \hat{a}_l^\dagger \hat{a}_k \int \psi_l^* U(\xi) \psi_k d\xi. \quad (99.23)$$

In the case of a system of particles between which there is a pair interaction the interaction energy operator has the form $\frac{1}{2} \sum_{i \neq j} W(\xi_i, \xi_j)$. Making use of (99.17) we write the Hamiltonian in the second-quantization representation

$$\hat{H} = \sum_{k,l} (\hat{H}_i)_{lk} \hat{a}_l^\dagger \hat{a}_k + \frac{1}{2} \sum_{k,p,l,m} \langle lm | W | kp \rangle \hat{a}_l^\dagger \hat{a}_m^\dagger \hat{a}_k \hat{a}_p, \quad (99.24)$$

or, taking as the functions ψ_k the eigenfunctions of the operator \hat{H}_i

$$\hat{H} = \sum_k E_k \hat{a}_k^\dagger \hat{a}_k + \frac{1}{2} \sum_{k,p,l,m} \langle lm | W | kp \rangle \hat{a}_l^\dagger \hat{a}_m^\dagger \hat{a}_k \hat{a}_p. \quad (99.25)$$

We note that the pair interaction (the last term of formula (99.24)) has an obvious interpretation. The interaction can be treated as the collision of two particles which are in the p th and k th states. After the interaction they make a transition to the l th and m th states.

It is useful to note that formula (99.20) can be obtained by means of the following formal method. In the expression for the mean energy (65.9) we replace the wave function by the operator in the space of occupation numbers defined as

$$\psi(\xi) \rightarrow \hat{\psi}(\xi) = \sum_k \hat{a}_k \psi_k(\xi) \quad (99.26)$$

and correspondingly

$$\psi^*(\xi) \rightarrow \hat{\psi}^\dagger(\xi) = \sum_l \hat{a}_l^\dagger \psi_l^*(\xi). \quad (99.26')$$

Then on the right-hand side of (65.9) we have

$$\int \psi^*(\xi) \hat{H}_i \psi(\xi) d\xi \rightarrow \sum_{k,l} \int \hat{a}_l^\dagger \psi_l^*(\xi) \hat{H}_i \hat{a}_k \psi_k(\xi) d\xi = \sum_{k,l} \hat{a}_l^\dagger \hat{a}_k (\hat{H}_i)_{lk}. \quad (99.27)$$

Comparing (99.27) and (99.20) we see that when the ordinary wave function is replaced by the operator the right-hand side of (65.9) is the same as (99.20). This means that in this case \bar{H} can formally be replaced by the operator \hat{H} in the second-quantization representation.

The name second quantization is due to the replacement of the wave function ψ by the operator $\hat{\psi}$. In second quantization not only are all mechanical quantities replaced by quantum operators (ordinary quantization) but the wave function itself is also quantized, i.e. replaced by an operator. Although second quantization is a formal method, it turns out to be very useful in a number of cases.

The Hamiltonian of a system of particles interacting in pairs can also be obtained easily in an analogous way. For this we again replace the functions ψ and ψ^* in formula (65.8) by operators (99.26). Then, in correspondence with what was said above, we make the replacement $\bar{H} \rightarrow \hat{H}$, where \hat{H} is the Hamiltonian in the second-quantization representation.

After the replacement we obtain formula (99.24).

All the results obtained so far apply to bosons. It can be shown* that formulae (99.20) and (99.24) remain valid also for a system of fermions. However, the operators \hat{a}_k and \hat{a}_k^\dagger can then no longer satisfy relations (99.9). Indeed, for the operators \hat{a}_k and \hat{a}_k^\dagger defined by formulae (99.9) the eigenvalues of the product $\hat{a}_k^\dagger \hat{a}_k$ are equal to arbitrary positive integers n_k . For a system of fermions the occupation numbers can be equal only to zero or one in accordance with the Pauli principle. The operators \hat{a}_k and \hat{a}_k^\dagger must now be defined in such a way that the eigenvalues of the operator $\hat{a}_k^\dagger \hat{a}_k$ are equal either to zero or to one, i.e.

$$(\hat{a}_k^\dagger \hat{a}_k)_{n_k n_k} = n_k = \begin{cases} 0 \\ 1 \end{cases}. \quad (99.28)$$

We shall show that conditions (99.28) are fulfilled if the operators \hat{a}_k and \hat{a}_k^\dagger satisfy the following anticommutation rules:

$$\hat{a}_k \hat{a}_l^\dagger + \hat{a}_l^\dagger \hat{a}_k = \delta_{kl}, \quad (99.29)$$

* See L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

$$\hat{a}_k \hat{a}_l + \hat{a}_l \hat{a}_k = \hat{a}_k^\dagger \hat{a}_l^\dagger + \hat{a}_l^\dagger \hat{a}_k^\dagger = 0. \quad (99.30)$$

For this we convince ourselves of the fact that

$$(\hat{a}_k^\dagger \hat{a}_k)^2 = \hat{a}_k^\dagger \hat{a}_k. \quad (99.31)$$

Indeed, we evaluate the left-hand side and making use of (99.29) we obtain

$$(\hat{a}_k^\dagger \hat{a}_k)^2 = \hat{a}_k^\dagger \hat{a}_k \hat{a}_k^\dagger \hat{a}_k = \hat{a}_k^\dagger \hat{a}_k (1 - \hat{a}_k \hat{a}_k^\dagger) = \hat{a}_k^\dagger \hat{a}_k - \hat{a}_k^\dagger \hat{a}_k \hat{a}_k \hat{a}_k^\dagger = \hat{a}_k^\dagger \hat{a}_k,$$

since $\hat{a}_k^2 = 0$, which follows from (99.30).

Taking the diagonal matrix elements of relation (99.31), we find $n_k^2 = n_k$. This equality can be fulfilled only for $n_k = 0$ and $n_k = 1$. One can find the explicit form of the matrices \hat{a}_k based on relations (99.30). Since the numbers n_k take on only two values 0 and 1, the operators \hat{a}_k and \hat{a}_k^\dagger are two-row matrices with respect to these variables. We shall present the corresponding matrix elements without derivation. They are

$$(a_k)_{01} = (a_k^\dagger)_{10} = \prod_{l=1}^{k-1} (1 - 2n_l). \quad (99.32)$$

All other matrix elements are equal to zero. As a result of the multiplication of the quantities $1 - 2n_l$, where $l = 1, 2, \dots, k-1$, either +1 or -1 is obtained, depending on the value of the occupation numbers of states preceding the given state.

Hence it is clear that the numbering of states 1, 2, ..., k , chosen initially, must not be changed.

The Schrödinger equation in the occupation number representation, where the Hamiltonian is given by formula (99.24), involves the law of conservation of the total number of particles (see §7). However, the introduction of the operators \hat{a}_k^\dagger and \hat{a}_k describing the absorption and production of particles allows one in a corresponding generalization also to investigate processes in which the number of particles of a given kind is not conserved.

§ 100. The quantum mechanics of the photon

The experimental establishment of the quantum or corpuscular nature of light was a spur to the creation of quantum theory as a whole.

On the other hand, the construction, as a consequence, of the quantum theory of the electromagnetic field has been one of the most notable successes of quantum theory.

Light quanta or photons are elementary particles whose distinctive property is the fact that their rest mass is equal to zero. Hence they always move with the velocity c in vacuum. This fact leads to certain important features of the method of describing their behaviour. Namely, the relation between the energy and momentum of the photon is given by the general formula

$$\epsilon = cp = \hbar ck. \quad (100.1)$$

If the momentum of the photon is replaced by its momentum operator, then the energy operator in the momentum representation has the form

$$\hat{H} = c\hat{p} = \hbar c\hat{k}. \quad (100.2)$$

Correspondingly the Schrödinger equation can be written in the momentum representation as

$$i\hbar \frac{\partial \psi_p}{\partial t} = \hat{H}\psi_p, \quad (100.3)$$

where ψ_p is the wave function of the photon in the momentum representation.

The operator \hat{H} is related to the photon energy ϵ by the general formula

$$\epsilon = \int \psi_p^* \hat{H} \psi_p \, d\mathbf{p} = \hbar c \int \psi_p^* \hat{k} \psi_p \, d\mathbf{p}. \quad (100.4)$$

On the other hand it can be assumed that there corresponds to a photon an electromagnetic field over all space. Its energy is

$$\epsilon = \int \frac{E^2 + H^2}{8\pi} \, dV = \frac{1}{4\pi} \int E^2 \, dV. \quad (100.5)$$

It is natural to identify the energy of the photon with the energy of the electromagnetic field. Both field vectors satisfy Maxwell's equations, which in a vacuum are reduced to the form

$$\nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0$$

and analogously for the vector \mathbf{H} .

Expanding \mathbf{E} in a Fourier integral

$$\mathbf{E}(\mathbf{r}, t) = \int \mathbf{E}(\mathbf{k}, t) e^{i\mathbf{k} \cdot \mathbf{r}} \, d\mathbf{k}$$

we have

$$\frac{\partial^2 \mathbf{E}(\mathbf{k}, t)}{\partial t^2} + k^2 \mathbf{E}(\mathbf{k}, t) = 0,$$

or

$$\left[\frac{\partial E(\mathbf{k}, t)}{\partial t} - i k E(\mathbf{k}, t) \right] \left[\frac{\partial E(\mathbf{k}, t)}{\partial t} + i k E(\mathbf{k}, t) \right] = 0. \quad (100.6)$$

By virtue of the fact that the field is real the following condition must be fulfilled:

$$E(\mathbf{k}) = E(-\mathbf{k}). \quad (100.7)$$

In place of the Fourier component $E(\mathbf{k}, t)$ we introduce the new function $f(\mathbf{k}, t)$ defined by the relations

$$\begin{aligned} E(\mathbf{k}, t) &= N(k) [f(\mathbf{k}, t) + f^*(-\mathbf{k}, t)], \\ \dot{E}(\mathbf{k}, t) &= -ikN(k) [f(\mathbf{k}, t) - f^*(-\mathbf{k}, t)] \end{aligned} \quad (100.8)$$

where N is a factor of proportionality. The dot denotes differentiation with respect to time.

It is easily seen that in such a representation of $E(\mathbf{k}, t)$ the condition (100.7) is automatically fulfilled.

Substituting the values of $E(\mathbf{k}, t)$ and $\dot{E}(\mathbf{k}, t)$ into (100.6) we arrive at two equations

$$i \frac{\partial f}{\partial t} = k f, \quad -i \frac{\partial f^*}{\partial t} = k f^*. \quad (100.9)$$

We stress that eqs. (100.9) only represent another form of notation of Maxwell's equations. Multiplying (100.9) by \hbar we obtain

$$i\hbar \frac{\partial f}{\partial t} = p f, \quad -i\hbar \frac{\partial f^*}{\partial t} = p f^*. \quad (100.10)$$

We see that the function $f(\mathbf{k}, t)$ satisfies an equation which is in form identical with the Schrödinger equation. If p is replaced by the operator \hat{H} , then the function $f(\mathbf{k}, t)$ must be identified with the wave function of the photon in the \mathbf{k} -representation.

The factor of proportionality N , which has so far remained arbitrary, can be defined from the comparison of (100.4) and (100.5).

Substituting expressions (100.8) into (100.5), we have

$$\begin{aligned}
\epsilon &= \frac{1}{4\pi} \int \mathbf{E}(\mathbf{k}, t) \cdot \mathbf{E}(\mathbf{k}', t) e^{i(\mathbf{k}+\mathbf{k}') \cdot \mathbf{r}} d\mathbf{k} d\mathbf{k}' dV = \\
&= \frac{1}{4\pi} \int \mathbf{E}(\mathbf{k}, t) \cdot \mathbf{E}(\mathbf{k}', t) d\mathbf{k} d\mathbf{k}' \int e^{i(\mathbf{k}+\mathbf{k}') \cdot \mathbf{r}} dV = \\
&= \frac{(2\pi)^3}{4\pi} \int \mathbf{E}(\mathbf{k}, t) \cdot \mathbf{E}(\mathbf{k}', t) \delta(\mathbf{k}+\mathbf{k}') d\mathbf{k} d\mathbf{k}' = \\
&= 2\pi^2 \int \mathbf{E}(\mathbf{k}, t) \cdot \mathbf{E}(-\mathbf{k}, t) d\mathbf{k} = 4\pi^2 \int N^2(\mathbf{k}) \mathbf{f}(\mathbf{k}) \cdot \mathbf{f}^*(\mathbf{k}) d\mathbf{k}.
\end{aligned}$$

For $N = (ck/4\pi^2)^{\frac{1}{2}}$ the energy of the electromagnetic field and the energy of the photon turn out to be identical. Thus in the \mathbf{k} -representation the photon is described by the wave function

$$\psi(\mathbf{k}, t) = \mathbf{f}(\mathbf{k}, t).$$

Then the following condition is fulfilled:

$$\int \mathbf{f}^* \mathbf{f} d\mathbf{k} = 1.$$

In this case the Maxwell equations for the electromagnetic field of a monochromatic wave turn out to be identical with the Schrödinger equation for an individual photon. Introducing the explicit dependence on time, we can write

$$\psi(\mathbf{k}, t) = \mathbf{f}_0(\mathbf{k}) e^{-i\omega t} = \mathbf{f}_0(\mathbf{k}) e^{-(i/\hbar)\epsilon t}.$$

By virtue of Maxwell's equation $\nabla \cdot \mathbf{E} = 0$, the amplitude in \mathbf{k} -space satisfies the condition $\mathbf{k} \mathbf{f}_0(\mathbf{k}) = 0$. We shall not dwell on the problems of normalization of the wave function and on the calculation of other quantum-mechanical quantities of photons, for example spin angular momentum, parity and so on: we refer the reader to the monograph of Akhiezer and Berestetskii*.

We confine ourselves only to some remarks of theoretical importance. We stress, first of all, that since Maxwell's equations are relativistically invariant so is the Schrödinger equation for the photon.

This is natural, since the photon always moves with the velocity of light.

We have found the wave function of the photon in the \mathbf{k} -representation (or, what is the same, in the \mathbf{p} -representation). This wave function has the usual probabilistic meaning. However, the wave function of the photon in the \mathbf{x} -representation, which would allow one to establish the probability of localization of the photon at a given point of space, does not exist.

* A.I. Akhiezer and V.B. Berestetskii, *Quantum electrodynamics* (Interscience Publishers, New York, 1965).

For free particles of rest mass m_0 different from zero the wave function in the x -representation is obtained from the wave function in the p -representation by means of the Fourier transformation.

In our case the Fourier transformation gives

$$\mathbf{f}(\mathbf{r}, t) = \int \mathbf{f}(\mathbf{k}, t) e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k}.$$

However, and here lies the fundamental difference between photons and particles with $m_0 \neq 0$, the position of a photon can be determined only as a result of interaction with charged particles, for example with electrons.

This interaction is determined by the value of the field vectors \mathbf{E} and \mathbf{H} at the point at which the electron is localized. The strength of the field at a certain point is defined by the Fourier inversion transformation, i.e.

$$\mathbf{E}(\mathbf{r}, t) = \int \mathbf{E}(\mathbf{k}, t) e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k} = \frac{1}{2\pi} \int (ck)^{\frac{1}{2}} [\mathbf{f}(\mathbf{k}, t) + \mathbf{f}^*(\mathbf{k}, t)] e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k}.$$

This formula shows that the field strength is not expressed in terms of $\mathbf{f}(\mathbf{r}, t)$, i.e. is not determined by the value of any wave function at the same point of space. On the contrary, $\mathbf{E}(\mathbf{r}, t)$ is determined by the distribution of $\mathbf{f}(\mathbf{r}, t)$ in space.

Photons have a spin equal to one. However, the definition of spin as the intrinsic angular momentum of the particle at rest makes no sense in the case of photons. Hence the division of the total angular momentum of the photon into an orbital part and a spin part is to a certain degree arbitrary.

This important last remark is associated with the description of a system of photons.

Photons do not interact directly with each other. The very weak interaction existing between photons is due to their interaction with electrons of the background. Hence the wave function of a system of photons is the wave function of a system of non-interacting particles. Photons as particles with integer spin obey Bose-Einstein statistics.

When photons interact with other particles the number of photons changes in the processes of emission and absorption. Photons are absorbed and emitted one at a time. The interaction of photons with charges can be described by means of their wave function (see the monograph of Akhiezer and Berestetskii cited above). However, this interaction is described in a much more effective and simple way by means of the second-quantization representation. We note that the method of second quantization was itself devised by Dirac for just this purpose.

§101. The quantization of the radiation field

As is well known, the development of quantum theory began with the establishment of the quantum properties of the electromagnetic field and the creation of a semiempirical theory of light quanta. Hence it is natural to try to apply the mathematical apparatus of quantum mechanics to the electromagnetic field. It turns out, however, that the electromagnetic field has a number of features which make this a complex problem. The modern quantum theory of the electromagnetic field, commenced by the studies of Dirac, is based on special methods, in particular on the method of second quantization*.

We recall that in the classical theory of the electromagnetic field in vacuum it was shown that a charge-free electromagnetic field can formally be compared to a mechanical system with an infinitely large number of degrees of freedom.

Expanding the vector potential, \mathbf{A} , of the electromagnetic field in terms of plane waves and taking the infinite set of amplitudes of the expansion q_i as generalized coordinates, it was possible to compare the electromagnetic field with a certain mechanical system; a set of field oscillators (see §38 of Part I). To each of the Fourier components of the expansion \mathbf{A} there corresponds one of the oscillators. Hence the complete set of field oscillators includes an infinitely large number and, consequently, the electromagnetic field could be compared to a mechanical system with an infinitely large number of degrees of freedom.

We write the Hamiltonian of this system as follows:

$$H = \sum \frac{1}{2} (p_\lambda^2 + \omega_\lambda^2 q_\lambda^2) = \sum H_\lambda, \quad (101.1)$$

where H_λ is the Hamiltonian of the λ th oscillator, p_λ is the generalized momentum corresponding to the coordinate q_λ , and ω_λ is the corresponding frequency. The summation is carried out over all values of frequencies and polarizations.

The quantum theory of the electromagnetic field is based on the assumption that this analogy can be given a direct physical content. Namely, it is assumed that a real electromagnetic field represents a quantum system which obeys the ordinary laws of quantum mechanics. The Hamiltonian H is ob-

* A more detailed exposition of the quantum theory of radiation may be found in the book of W. Heitler, *The quantum theory of radiation* (Clarendon Press, Oxford, 1954).

tained from the classical Hamiltonian (101.1) by means of the usual replacement of mechanical quantities, generalized coordinates and momenta, by corresponding quantum operators. That is, we replace q_λ and p_λ by operators satisfying the commutation relations:

$$\hat{p}_\lambda \hat{q}_\mu - \hat{q}_\mu \hat{p}_\lambda = \frac{\hbar}{i} \delta_{\lambda\mu}, \quad \hat{q}_\lambda \hat{q}_\mu - \hat{q}_\mu \hat{q}_\lambda = 0, \quad \hat{p}_\lambda \hat{p}_\mu - \hat{p}_\mu \hat{p}_\lambda = 0.$$

Since different field oscillators are independent, the operators \hat{p}_λ and \hat{q}_λ referring to different oscillators commute with each other. Then \hat{H} will represent the Hamiltonian of a quantum system. It is advisable, however, to carry out the canonical transformation to new variables (see formulae (50.11)). Namely, we write

$$\begin{aligned} \hat{a}_\lambda &= \frac{1}{\sqrt{2}} \left[\left(\frac{\omega_\lambda}{\hbar} \right)^{\frac{1}{2}} \hat{q}_\lambda + \frac{i\hat{p}_\lambda}{(\omega_\lambda \hbar)^{\frac{1}{2}}} \right], \\ \hat{a}_\lambda^\dagger &= \frac{1}{\sqrt{2}} \left[\left(\frac{\omega_\lambda}{\hbar} \right)^{\frac{1}{2}} \hat{q}_\lambda - \frac{i\hat{p}_\lambda}{(\omega_\lambda \hbar)^{\frac{1}{2}}} \right]. \end{aligned} \quad (101.2)$$

In the new representation

$$\hat{p}_\lambda^2 + \omega_\lambda^2 \hat{q}_\lambda^2 = \hbar \omega_\lambda (\hat{a}_\lambda \hat{a}_\lambda^\dagger + \hat{a}_\lambda^\dagger \hat{a}_\lambda),$$

so that

$$\hat{H} = \frac{1}{2} \sum_\lambda \hbar \omega_\lambda (\hat{a}_\lambda \hat{a}_\lambda^\dagger + \hat{a}_\lambda^\dagger \hat{a}_\lambda).$$

To the operators \hat{a}_λ and \hat{a}_λ^\dagger there correspond the commutation relations

$$\begin{aligned} \hat{a}_\lambda \hat{a}_\mu^\dagger - \hat{a}_\mu^\dagger \hat{a}_\lambda &= \delta_{\lambda\mu}, \\ \hat{a}_\lambda \hat{a}_\mu - \hat{a}_\mu \hat{a}_\lambda &= 0, \\ \hat{a}_\lambda^\dagger \hat{a}_\mu^\dagger - \hat{a}_\mu^\dagger \hat{a}_\lambda^\dagger &= 0, \end{aligned} \quad (101.3)$$

which follow immediately from the definition and commutation relations for \hat{p}_λ and \hat{q}_λ .

The Hamiltonian can be transformed by means of (101.3), writing

$$\hat{a}_\lambda \hat{a}_\lambda^\dagger = 1 + \hat{a}_\lambda^\dagger \hat{a}_\lambda.$$

Then

$$\hat{H} = \sum \hbar \omega_\lambda (\hat{a}_\lambda^\dagger \hat{a}_\lambda + \frac{1}{2}). \quad (101.4)$$

Comparing the expression (101.4) for \hat{H} and the commutation relations (101.3) for the operators \hat{a} and \hat{a}^\dagger with the corresponding expressions (99.9) and (99.21), we see that they are completely analogous. This means that a

free electromagnetic field represents a system of bosons which are usually called photons or light quanta.

To each plane wave in expansion (38.19) of Part I there corresponds a photon. The energy of each photon, according to formula (101.4), is equal to $\hbar\omega_\lambda$. The total energy of the electromagnetic field correspondingly has the form

$$E = \sum E_\lambda n_\lambda + \sum \frac{1}{2} \hbar \omega_\lambda = \sum E_\lambda n_\lambda + E_0, \quad (101.5)$$

where $E_\lambda = \hbar\omega_\lambda$, and n_λ is the number of photons with energy E_λ .

The second term of formula (101.5), denoted by E_0 , is called the energy of the zero-point oscillations of the electromagnetic field. Formula (101.5) shows that if all $n_\lambda = 0$, i.e. if there are no photons in the field, then the energy of the electromagnetic field is equal to E_0 . Moreover, the quantity E_0 itself is infinitely large, since the sum for E_0 involves an infinitely large number of positive terms $\hbar\omega_\lambda$.

The presence of the infinitely large constant term in the energy of the electromagnetic field has no effect on the processes of interaction of the field with matter (the emission, absorption and scattering of light) which will be considered in this chapter. In these processes changes occur in the state of the electromagnetic field for which only the difference between the energies of two states is important.

When an energy difference is formed the zero-point energy is cancelled. Hence, until recently, it was assumed that the zero-point energy could be taken as the zero of energy and that it could be formally omitted in all expressions. However, as quantum electrodynamics developed it turned out that this is not so and that the appearance of the term E_0 in the formula for the energy of the electromagnetic field has a profound meaning.

From the point of view of modern electrodynamics the 'emptiness', the absence of particles and of photons, is not 'nothing' but is a definite state of the field, called the vacuum. The existence of the vacuum state and of zero-point oscillations with frequencies ω_λ is important in certain interactions between the electromagnetic field and electrons and leads to a number of observed effects.

We shall touch briefly upon the problem of the vacuum in §116 and §128. In the meanwhile we shall not consider the zero-point energy.

Let us now find the momentum of a charge-free electromagnetic field. According to (38.25) of Part I, we have for the momentum of a plane wave

$$\mathbf{p}_\lambda = \frac{\mathbf{k}_\lambda}{k_\lambda c} E_\lambda, \quad (101.6)$$

where \mathbf{k}_λ and E_λ are respectively the wave vector and the energy of the wave.

If we pass to quantized expressions and replace E_λ by its eigenvalue, then we easily obtain

$$\mathbf{p}_\lambda = \hbar \mathbf{k}_\lambda.$$

Just as $\hbar\omega_\lambda$ represents the energy of an individual photon, $\hbar\mathbf{k}_\lambda$ is its momentum. We see that between the energy and momentum of the photon there is the relation found from the analysis of experimental data even before the creation of quantum mechanics

$$|\mathbf{p}_\lambda| = \frac{E_\lambda}{c}.$$

From (101.6) it follows, in particular, that the rest mass of the photon is equal to zero (see §14 of Part II). The total momentum of the electromagnetic field is equal to

$$\mathbf{P} = \sum \hbar \mathbf{k}_\lambda n_\lambda. \quad (101.7)$$

It is determined by the occupation numbers n_λ .

We now turn to the formulation of the Schrödinger equation for the electromagnetic field. It has the usual form

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H} \psi.$$

The wave function of the electromagnetic field is usually called the amplitude of the state of the field. If use is made of the Hamiltonian in the representation of occupation numbers, then the amplitude of the state of the electromagnetic field will also be a function of the occupation numbers n_λ

$$\psi = \psi(n_1, n_2, \dots, n_\lambda, \dots, t).$$

According to the conclusions of §99, the operators \hat{a}_λ^\dagger and \hat{a}_λ represent the photon creation and annihilation operators. When they act on the wave function they respectively increase and reduce by one the number of photons of frequency ω_λ . The matrix elements of these operators are given by formulae (99.6) and (99.7).

§102. The interaction of an electron with radiation

Having carried out the quantization of a free electromagnetic field, we can turn to the consideration of a system consisting of an electromagnetic field

and particles. We shall assume that there is one electron in the radiation field and shall find the interaction between the electron and the electromagnetic field. In this chapter we shall suppose that the electron has a velocity small in comparison with the velocity of light and that it is described by a non-relativistic Hamiltonian. We write the Hamiltonian of the system (radiation field + electron) in the form

$$\hat{H} = \frac{1}{2m} \left(\hat{\mathbf{p}} - \frac{e}{c} \hat{\mathbf{A}} \right)^2 + \hat{H}_{\text{rad}} .$$

We assume that the scalar potential φ is chosen to be equal to zero, and that the gauge condition (see (10.5) of Part I) of the vector potential \mathbf{A} is of the form $\nabla \cdot \mathbf{A} = 0$. From this relation it follows that the momentum operator $\hat{\mathbf{p}}$ commutes with the vector \mathbf{A} , and hence the Hamiltonian \hat{H} can be rewritten as

$$\hat{H} = \frac{\hat{\mathbf{p}}^2}{2m} - \frac{e}{mc} (\hat{\mathbf{p}} \cdot \hat{\mathbf{A}}) + \frac{e^2}{2mc^2} \hat{\mathbf{A}}^2 + \hat{H}_{\text{rad}} . \quad (102.1)$$

The first term of (102.1) represents the Hamiltonian of the free particle, and the last term the Hamiltonian of the free radiation field. The Hamiltonian of the interaction of the electron with the radiation field, responsible for all processes of emission and absorption of photons by the electron, is of the form

$$\hat{H}' = - \frac{e}{mc} (\hat{\mathbf{p}} \cdot \hat{\mathbf{A}}) + \frac{e^2}{2mc^2} \hat{\mathbf{A}}^2 . \quad (102.2)$$

We shall formally assume the electron charge to be the small parameter in terms of which the perturbation theory expansion is carried out. In what follows we shall in fact see that the expansion is carried out in powers of the small quantity $e^2/\hbar c = \frac{1}{137}$ which figures in the corresponding matrix elements and is called the interaction constant. We shall confine ourselves to the consideration of some simple processes in the first non-vanishing approximation of perturbation theory. We have, in §56, obtained the general expressions for the probabilities of different processes, and our problem reduces to the calculation of the matrix elements of the interaction operator \hat{H}' considered as the perturbation operator. The expansion of the vector potential is conveniently written in the form (38.19) of Part I:

$$\mathbf{A} = \sum_{\lambda} (b_{\lambda} \mathbf{A}_{\lambda} + b_{\lambda}^* \mathbf{A}_{\lambda}^*) ,$$

where

$$\mathbf{A}_\lambda = \mathbf{e}_\lambda (4\pi c^2/V)^{1/2} e^{i\mathbf{k}_\lambda \cdot \mathbf{r}}. \quad (102.3)$$

We pass to the quantum operators

$$\hat{\mathbf{A}} = \sum_{\lambda} (\hat{b}_\lambda \mathbf{A}_\lambda + \hat{b}_\lambda^\dagger \mathbf{A}_\lambda^*). \quad (102.3')$$

Making use of relations (38.20) of Part I, we express the operators \hat{b}_λ and \hat{b}_λ^\dagger in terms of the operators \hat{q}_λ and \hat{p}_λ

$$\hat{b}_\lambda = \frac{1}{2\omega_\lambda} (\omega_\lambda \hat{q}_\lambda + i\hat{p}_\lambda), \quad \hat{b}_\lambda^\dagger = \frac{1}{2\omega_\lambda} (\omega_\lambda \hat{q}_\lambda - i\hat{p}_\lambda).$$

Using formulae (101.2) we introduce the operators \hat{a}_λ and \hat{a}_λ^\dagger . We then obtain

$$\hat{b}_\lambda = (\hbar/2\omega_\lambda)^{1/2} \hat{a}_\lambda, \quad \hat{b}_\lambda^\dagger = (\hbar/2\omega_\lambda)^{1/2} \hat{a}_\lambda^\dagger. \quad (102.4)$$

Comparing with (99.6) and (99.7) we find that the operators \hat{b}_λ and \hat{b}_λ^\dagger have the following matrix elements different from zero:

$$\begin{aligned} \langle n_1, \dots, n_\lambda, \dots | \hat{b}_\lambda | n_1, \dots, n_\lambda + 1, \dots \rangle &= [\hbar(n_\lambda + 1)/2\omega_\lambda]^{1/2}, \\ \langle n_1, \dots, n_\lambda, \dots | \hat{b}_\lambda^\dagger | n_1, \dots, n_\lambda - 1, \dots \rangle &= [\hbar n_\lambda / 2\omega_\lambda]^{1/2}. \end{aligned} \quad (102.5)$$

Thus the matrix elements of the vector potential differ from zero only for the processes of emission and absorption of one photon. For the operator $\hat{\mathbf{A}}^2$ involved in (102.2) we have

$$\hat{\mathbf{A}}^2 = \sum_{\lambda, \lambda'} [\hat{b}_\lambda \hat{b}_{\lambda'} (\mathbf{A}_\lambda \mathbf{A}_{\lambda'}) + \hat{b}_\lambda \hat{b}_{\lambda'}^\dagger (\mathbf{A}_\lambda \mathbf{A}_{\lambda'}^*) + \hat{b}_\lambda^\dagger \hat{b}_{\lambda'} (\mathbf{A}_\lambda^* \mathbf{A}_{\lambda'}) + \hat{b}_\lambda^\dagger \hat{b}_{\lambda'}^\dagger (\mathbf{A}_\lambda^* \mathbf{A}_{\lambda'}^*)]. \quad (102.6)$$

From this expression it is seen that the matrix elements of the operator $\hat{\mathbf{A}}^2$ differ from zero for two-photon transitions, i.e. for the emission or absorption of two photons or the emission of one photon and the absorption of another.

The term containing the operator $\hat{\mathbf{A}}^2$ as well as the term with operator $-(e/mc)(\hat{\mathbf{p}} \cdot \hat{\mathbf{A}})$ gives a contribution to processes involving two photons, the latter term being taken into account in the second approximation of perturbation theory.

The vector potential (102.3) describes the state of a photon with given

momentum. One can also introduce the concept of the state of a photon with given angular momentum. In order to find the expression for the vector potential describing the state of a photon with angular momentum and its component along the z -axis we should carry out the expansion of the vector potential \mathbf{A} not in terms of plane waves, but in terms of spherical waves. The amplitudes of the expansion must be considered as operators in the space of occupation numbers satisfying commutation relations of the same type as (101.3). In a state with given momentum the angular momentum of the photon does not have a definite value. This corresponds to the fact that the plane wave can be written in the form of an expansion in terms of an infinite sequence of spherical waves.

The photon possesses definite 'internal' degrees of freedom, since in describing its state it is necessary to take into account different possible polarizations.

The 'internal' state of a system is usually associated with its spin. However, the definition of the spin of a system as its 'intrinsic' angular momentum, i.e. the angular momentum at rest, is inapplicable to the photon. The photon in any reference frame moves with velocity c .

Nevertheless, it sometimes appears to be convenient to introduce the concept of spin also for the photon, writing the total angular momentum operator in the form of a superposition of the orbital angular momentum operator and the spin operator. In this case it turns out that the spin of the photon must be considered to be equal to one. In correspondence with three possible spin components $s_z = 0, \pm 1$, one would think that the photon may be in three different states with different polarization. However, the condition of the transverse nature of electromagnetic waves leads to the fact that actually only two spin components are possible, which correspond to the two independent polarization states of the photon. The reader may find a detailed consideration of the problems which are touched upon here in the monograph of Akhiezer and Berestetskii*.

§103. The absorption and emission of light

Let us consider the probability of a one-photon transition; i.e. the process of absorption and emission. We shall first of all write down the matrix elements corresponding to the absorption and emission of a photon of frequency

* A.I. Akhiezer and V.B. Berestetskii, *Quantum electrodynamics* (Interscience Publishers, New York, 1965).

ω_λ . Suppose the electron was in the initial state ψ_1 before absorption and in the state ψ_2 after absorption. The transition $1 \rightarrow 2$ proceeds with the absorption, and the transition $2 \rightarrow 1$ with the emission of a photon of frequency ω_λ . The matrix element of the perturbation operator (102.2) for the transition with the absorption of a photon is of the form

$$\begin{aligned} \langle 2, n_\lambda - 1 | \hat{H}' | 1, n_\lambda \rangle &= \\ &= -\frac{e}{mc} \int \psi_2^* (\hat{\mathbf{p}} \cdot \mathbf{e}_\lambda) \left(\frac{4\pi c^2}{V} \right)^{\frac{1}{2}} e^{i\mathbf{k}_\lambda \cdot \mathbf{r}} (b_\lambda)_{n_\lambda - 1, n_\lambda} \psi_1 dV = \\ &= -\frac{e}{m} \left(\frac{2\pi\hbar n_\lambda}{V\omega_\lambda} \right)^{\frac{1}{2}} \int \psi_2^* (\hat{\mathbf{p}} \cdot \mathbf{e}_\lambda) e^{i\mathbf{k}_\lambda \cdot \mathbf{r}} \psi_1 dV. \end{aligned} \quad (103.1)$$

Analogously for the process of emission of a photon we have

$$\langle 1, n_\lambda + 1 | \hat{H}' | 2, n_\lambda \rangle = -\frac{e}{m} \left(\frac{2\pi\hbar(n_\lambda + 1)}{V\omega_\lambda} \right)^{\frac{1}{2}} \int \psi_1^* (\hat{\mathbf{p}} \cdot \mathbf{e}_\lambda) e^{-i\mathbf{k}_\lambda \cdot \mathbf{r}} \psi_2 dV. \quad (103.2)$$

The probability per unit time of the transition with the absorption of a photon is given by the formula (see §56)

$$dW = \frac{2\pi}{\hbar} |\langle 2, n_\lambda - 1 | \hat{H}' | 1, n_\lambda \rangle|^2 \rho(\omega) d\Omega. \quad (103.3)$$

Here $d\Omega$ is the solid angle element corresponding to the direction of propagation of the photon before absorption. We shall assume that states 1 and 2 of the electron belong to a discrete spectrum. In this case the final state of the system with energy E_2 belongs to a discrete spectrum, while the initial state with energy $E_1 + \hbar\omega$ belongs to a continuous spectrum (since the frequency ω changes in a continuous way). Then the photon absorbed may belong to any of the oscillators in the interval of states $\omega d\Omega$ in volume V . The number of such oscillators for given polarization per unit volume is given by formula (38.23) of Part I. Passing to a continuous distribution of frequencies we shall omit the index λ where this cannot lead to misunderstanding, or replace it by the index \mathbf{k} .

By $\rho(\omega)$ in expression (103.3) is meant the number of oscillators in volume V corresponding to unit energy and angular intervals for a given polarization:

$$\rho(\omega) = \frac{\omega^2 V}{(2\pi c)^3 \hbar}. \quad (103.4)$$

For the transition probability per unit time taking into account (103.1) we obtain

$$dW = \frac{e^2 \omega}{m^2 2\pi \hbar c^3} |((\hat{\mathbf{p}} \cdot \mathbf{e}) e^{i\mathbf{k} \cdot \mathbf{r}})_{21}|^2 n_{\mathbf{k}} d\Omega. \quad (103.5)$$

The absorption probability is equal to zero for all energies except those which satisfy the conservation law

$$E_2 = E_1 + \hbar \omega. \quad (103.6)$$

Let us determine the intensity $J_0(\omega)$ of the incident radiation corresponding to the frequency interval $d\omega$ and angular interval $d\Omega$. Since to one oscillator there correspond $n_{\mathbf{k}}$ photons with given polarization, we have

$$J_0(\omega) d\omega d\Omega = n_{\mathbf{k}} \hbar \omega c \rho \hbar d\omega d\Omega = n_{\mathbf{k}} \hbar \frac{\omega^3 d\omega d\Omega}{(2\pi)^3 c^2}.$$

The total probability is proportional to the intensity of the incident radiation. The probability of emission of a photon by the electron is easily calculated in a completely analogous way.

The probability per unit time of the transition with the emission of a photon with momentum $\hbar \mathbf{k}$ and polarization \mathbf{e} is given by a formula of the type of (103.3)

$$dW = \frac{e^2 \omega}{m^2 2\pi \hbar c^3} |((\hat{\mathbf{p}} \cdot \mathbf{e}) e^{-i\mathbf{k} \cdot \mathbf{r}})_{12}|^2 (n_{\mathbf{k}} + 1) d\Omega. \quad (103.7)$$

The emission probability is different from zero if the frequency of the emitted quantum is equal to

$$\hbar \omega = E_2 - E_1. \quad (103.8)$$

We see further that the probability of the transition $2 \rightarrow 1$ with the emission of a photon, given by formula (103.7), consists of two terms. One of these is proportional to the intensity of radiation (to the number of photons $n_{\mathbf{k}}$) existing before the emission. The initially existing electromagnetic field acts on the electron, favouring its transition into a new state with the emission of an additional photon. This is called stimulated emission. The existence of stimulated emission was first pointed out by Einstein before the creation of the modern quantum theory of radiation. The second term of formula (103.7) does not depend on the intensity of the initial radiation and also ensures the possibility of emission in the case where before the emission the

electromagnetic field was not excited (the number of photons $n_k = 0$). Emission of such a type is called spontaneous emission.

From the comparison of formulae (103.5) and (103.7), taking into account the hermitian property of the matrix elements, it follows that for the ratio of the probabilities of emission and absorption of a photon one can write

$$\frac{dW_{\text{emiss}}}{dW_{\text{abs}}} = \frac{n_k + 1}{n_k}. \quad (103.9)$$

We shall see in §12 of Part VI (Volume 4) that it is easy to obtain from (103.9) the Planck formula for the intensity distribution in black-body radiation.

We shall now show that only electrons in bound states can absorb and emit photons. For this we calculate the integrals involved in the matrix elements for the transition probabilities, assuming the electron to be free. The wave functions ψ_1 and ψ_2 are written in the form of plane waves

$$\psi_1 = C e^{(i/\hbar)(\mathbf{p}_1 \cdot \mathbf{r})}, \quad \psi_2 = C e^{(i/\hbar)(\mathbf{p}_2 \cdot \mathbf{r})},$$

where C is the normalization constant. Substituting these wave functions into (103.2), we easily find

$$\begin{aligned} \int \psi_1^* \left(\frac{\hbar}{i} \nabla \cdot \mathbf{e} \right) e^{-i\mathbf{k} \cdot \mathbf{r}} \psi_2 dV = \\ = |C|^2 \int e^{(-i/\hbar)(\mathbf{p}_1 \cdot \mathbf{r})} \left(\frac{\hbar}{i} \mathbf{e} \cdot \nabla \right) e^{(i/\hbar)(\mathbf{p}_2 - \hbar \mathbf{k}) \cdot \mathbf{r}} dV \sim \delta(\mathbf{p}_2 - \hbar \mathbf{k} - \mathbf{p}_1). \end{aligned} \quad (103.10)$$

Formula (103.10) expresses the momentum conservation law in the interaction of a photon with a free electron. Furthermore, the energy conservation law holds in the transition. Thus the following equalities must be fulfilled simultaneously

$$\mathbf{p}_2 = \mathbf{p}_1 + \hbar \mathbf{k}, \quad (103.11)$$

$$E_2 = E_1 + \hbar \omega. \quad (103.12)$$

It is easily seen that eqs. (103.11) and (103.12) are inconsistent. An analogous conclusion, of course, also applies to the case of absorption.

For the laws of conservation of energy and momentum to hold simultaneously it is necessary that a third body, to which the excess momentum is transferred, be involved. In the case of atomic electrons such a body can be the nucleus of the atom.

§104. Dipole transitions in atomic systems

The matrix element for the process of emission of a photon (103.2) can in most cases be substantially simplified. Usually the wavelength of the photon emitted is considerably larger than the linear size of the region of space in which the wave functions of the electron ψ_1 and ψ_2 are considerably different from zero.

For example, let the electron move in an atom whose effective radius is equal to a . Then the wave functions of the initial and final states are very small outside the range a . The energy of the electron in the field of the nucleus with effective charge Z^* is in order of magnitude equal to Z^*e^2/a . The change ΔE in the energy of the atom in the transition and, consequently, the energy of the emitted photon is of the same order of magnitude. Then the length of the emitted wave is $\lambda \approx c/\omega \approx \hbar c/\hbar\omega \approx \hbar ca/Z^*e^2$. The ratio of the atomic size to the wavelength is of the order of

$$\frac{a}{\lambda} \approx \frac{Z^*e^2}{\hbar c} \approx \frac{Z^*}{137}.$$

For the outer electrons $Z^* \approx 1$ and the wavelength is substantially larger than the atomic size. In the case of X-radiation arising in transitions in the K-shell of heavy atoms this approximation turns out to be inadequate. For $\lambda \gg a$ the index of the exponential function inside the integral in (103.2) is very small within the limits of the effective range of integration, and hence the factor $e^{-ik \cdot r}$ can be replaced by unity.

The probability of the transition with emission, (103.7), is then rewritten in the form

$$dW = \frac{e^2\omega}{m^2 2\pi\hbar c^3} |(p_e)_{12}|^2 (n_k+1) d\Omega. \quad (104.1)$$

Here \hat{p}_e is the operator of the component of the momentum of the particle along the direction of polarization of the emitted quantum.

The matrix element of the momentum operator can be expressed in terms of the matrix element of the coordinate. According to (31.7) and (49.5) we have

$$p_{12} = m\mathbf{v}_{12} = m\dot{\mathbf{r}}_{12} = \frac{im}{\hbar} (E_1 - E_2) \mathbf{r}_{12} = -\frac{im}{e} \omega \mathbf{d}_{12}, \quad (104.2)$$

where \mathbf{d} is the dipole moment of the particle. Substituting (104.2) into (104.1) we obtain

$$dW = \frac{\omega^3}{2\pi\hbar c^3} |(d_e)_{12}|^2 (n_{\mathbf{k}}+1) d\Omega. \quad (104.3)$$

Here d_e is the component of the dipole moment vector of the particle along the direction of polarization. We see that the transition probability (104.3) depends on the matrix element of the dipole moment of the particle and hence such transitions are called dipole transitions and the radiation is called dipole radiation. If the angle between $(\mathbf{d})_{12}$ and the direction of polarization of the radiation is denoted by θ , then expression (104.3) can be rewritten as

$$dW = \frac{\omega^3}{2\pi\hbar c^3} |\mathbf{d}_{12}|^2 (n_{\mathbf{k}}+1) \cos^2 \theta d\Omega. \quad (104.4)$$

We sum this expression over the polarizations of the quantum. As independent directions of polarization we choose the polarization in the plane (\mathbf{d}, \mathbf{k}) and the polarization in the direction perpendicular to this plane. Expression (104.4) is then brought into the form

$$dW = \frac{\omega^3}{2\pi\hbar c^3} |\mathbf{d}_{12}|^2 (n_{\mathbf{k}}+1) \sin^2 \vartheta d\Omega, \quad (104.5)$$

where ϑ is the angle between the vector \mathbf{d}_{12} and the direction of propagation of the radiation \mathbf{k} .

The intensity of emission per unit time into the element of solid angle $d\Omega$ is obtained by multiplying (104.5) by the energy of the photon $\hbar\omega$. For spontaneous emission we have

$$J d\Omega = \frac{\omega^4}{2\pi c^3} |\mathbf{d}_{12}|^2 \sin^2 \vartheta d\Omega. \quad (104.6)$$

Integrating over angles we find the total spontaneous emission per unit time

$$\frac{dE}{dt} = \frac{4\omega^4}{3c^3} |\mathbf{d}_{12}|^2. \quad (104.7)$$

This expression is very similar to the classical formula for the intensity of dipole radiation (see (27.9) of Part I). The difference between the classical and quantum formulae lies only in the fact that the averaged square of the dipole moment $\overline{\mathbf{d}^2}$ involved in the classical expression must be replaced by the corresponding matrix element (doubled) $2|\mathbf{d}_{12}|^2$.

Dipole transitions in the absorption of light can be considered in an analogous way. Setting $e^{i\mathbf{k}\cdot\mathbf{r}} = 1$ in (103.1) and taking into account (104.2), we obtain for the transition probability per unit time

$$dW = \frac{\omega^3}{2\pi\hbar c^3} |\mathbf{d}_{21}|^2 n_{\mathbf{k}} \cos^2 \theta d\Omega. \quad (104.8)$$

Averaging this expression over all orientations of the vector \mathbf{d} with respect to the direction of incident radiation, we find

$$\overline{\cos^2 \theta} = \frac{1}{4\pi} \int \cos^2 \theta d\Omega = \frac{1}{3}. \quad (104.9)$$

Expressing $n_{\mathbf{k}}$ in terms of the intensity of incident radiation $J_0(\omega)$ and multiplying (104.8) by $\hbar\omega$, we find the energy absorbed per unit time

$$J d\Omega = \frac{4\pi^2 e^2}{3 \hbar c} \omega |\mathbf{r}_{21}|^2 J_0(\omega) d\Omega. \quad (104.10)$$

So far we have considered the absorption and emission of a photon by one electron. If the absorbing or emitting system contains several electrons, then, disregarding the interaction between them, it can be assumed that formulae (104.10) and (104.5) will remain valid, provided that the dipole moment of the electron in them is replaced by the sum of the dipole moments of all the electrons.

§105. Quadrupole and magnetic dipole radiation

The matrix elements of a dipole transition are obtained from the general expression (103.2) when the exponential function $e^{-i\mathbf{k}\cdot\mathbf{r}}$ is replaced by unity. It may turn out, however, that the matrix element of the dipole transition reduces to zero, whereas the precise matrix element (103.2) differs from zero. In this case one has to expand the exponential $e^{-i\mathbf{k}\cdot\mathbf{r}}$ in a series, writing out the higher terms of the expansion. Then the emission probability will be different from zero, although substantially lower than the probability of dipole radiation. For this reason such transitions are called forbidden. The emission probability determined by the following terms of the expansion will have the form

$$dW = \frac{e^2 \omega^3}{2\pi\hbar c^3} |(\mathbf{r}_e(\mathbf{k}\cdot\mathbf{r}))_{21}|^2 (n_{\mathbf{k}}+1) d\Omega. \quad (105.1)$$

The intensity of spontaneous emission in such a transition, analogously to (104.6), will be equal to

$$J d\Omega = \frac{e^2 \omega^4}{2\pi c^3} |(\mathbf{r}(\mathbf{k} \cdot \mathbf{r}))_{21}|^2 \sin^2 \vartheta d\Omega. \quad (105.2)$$

Comparing this formula with the classical expressions (see Vol. 1, §31 of Part I), we see that (105.2) represents the magnetic dipole and quadrupole radiation. The probability of the forbidden radiation (magnetic dipole and quadrupole) is related to the probability of allowed dipole radiation as a^2/λ^2 ($k \approx \lambda^{-1}$, $r \approx a$). If for some reason or other the matrix elements (105.1) are equal to zero, then the probability of radiation of higher order can be found in an analogous way.

§106. Selection rules

We see that the character of radiation from atomic and nuclear systems is determined by the matrix element $\mathbf{d}_{21} = e\mathbf{r}_{21}$. Let us now establish when this matrix element can be different from zero, i.e. between which states of the system transitions accompanied by dipole radiation are possible. The set of requirements which must be satisfied by the wave functions of the initial and final states of the system in order that the matrix element of the dipole transition \mathbf{r}_{21} may not reduce to zero are called the selection rules for dipole radiation. The selection rules can easily be formulated in the general form if the wave functions ψ_1 and ψ_2 describe the state of a particle moving in a centrally symmetric field. In this case the dependence of ψ_1 and ψ_2 on angles is characterized by spherical functions (see §35). For dipole transitions to be possible in the system the matrix element of the projection of the radius vector on the direction of polarization of the quantum e must be different from zero. Let us first consider a quantum polarized along the z -axis. In this case $r_e = z = r \cos \vartheta$. The matrix element of the dipole transition will be proportional to the integral

$$\int_0^\pi \int_0^{2\pi} Y_{l_2 m_2}^* \cos \vartheta Y_{l_1 m_1} \sin \vartheta d\vartheta d\varphi. \quad (106.1)$$

Here l_1 , m_1 and l_2 , m_2 are the quantum numbers of the states of the system before and after the emission of the quantum. Taking into account the definition of spherical functions (30.16), the integral (106.1) can be rewritten in the form

$$\int_0^\pi P_{l_2}^{m_2}(\cos \vartheta) P_{l_1}^{m_1}(\cos \vartheta) \cos \vartheta \sin \vartheta d\vartheta \int_0^{2\pi} e^{i(m_1 - m_2)\varphi} d\varphi. \quad (106.2)$$

The integral over the angle φ is different from zero only for $m_1 = m_2$. The integral over the angle ϑ then has the form

$$\int_{-1}^1 P_{l_2}^m(x) x P_{l_1}^m(x) dx. \quad (106.3)$$

It can be shown that the following relation is valid for associated Legendre polynomials*:

$$x P_l^m(x) = \frac{l+|m|}{2l+1} P_{l-1}^m(x) + \frac{l-|m|+1}{2l+1} P_{l+1}^m(x). \quad (106.4)$$

Substituting this expression into (106.3) and taking into account the conditions of orthogonality of associated Legendre polynomials, we find that integral (106.3) is different from zero only for $l_2 = l_1 \pm 1$.

Thus we see that if the radiation is polarized along the z-axis, then the matrix element of the dipole transition differs from zero only for transitions with $m_2 = m_1$, $l_2 = l_1 \pm 1$.

Let us now define analogous selection rules for the quantum numbers l, m in the case where the quantum is emitted in the direction of the z-axis and, consequently, is polarized in the (x, y) -plane. We consider the case of circular polarization with a phase shift equal to $\frac{1}{2}\pi$. Then the transition probability is determined by the matrix element of the quantity $x \pm iy$

$$(x \pm iy)_{21} = (r \sin \vartheta e^{\pm i\varphi})_{21}. \quad (106.5)$$

Separating the integral over the angle φ , we obtain

$$\int_0^{2\pi} e^{i(m_1 - m_2 \pm 1)\varphi} d\varphi. \quad (106.6)$$

This integral is different from zero under the condition

$$m_2 = m_1 \pm 1. \quad (106.7)$$

The corresponding integral over the angle ϑ is different from zero if $l_2 =$

* See, for example, N.N. Lebedev, *Special functions and their applications* (Prentice Hall, Englewood Cliffs, N.Y., 1965).

$l_1 \pm 1$. Thus the selection rules obtained for the quantum numbers l and m for dipole transitions can finally be formulated in the form

$$\Delta m = 0, \pm 1; \quad \Delta l = \pm 1. \quad (106.8)$$

It can easily be seen that the selection rules given by relations (106.8) express the angular momentum conservation law. The fact that l may change by one shows that in a dipole transition the emitted quantum carries away an angular momentum equal to one. At first sight this conclusion may seem to be strange. As a matter of fact, we have considered (formula (103.7)) transitions with the emission of a photon of given momentum. But in a state with given momentum and polarization the angular momentum of the photon does not have a sharp value. If, however, the wavelength of the photon is large in comparison with the size of the system, then it is possible to expand the function $e^{i\mathbf{k}\cdot\mathbf{r}}$ in a series. Carrying out this expansion and retaining the first non-vanishing term, i.e. the dominant term determining the value of the matrix element, we in fact separate photons with given total angular momentum. To dipole radiation there correspond photons with angular momentum one, to quadrupole radiation photons with angular momentum two and so on. Direct calculation, on which we cannot dwell, confirms this conclusion.

Selection rules (106.8) automatically satisfy the requirements of the parity conservation law. Since the operator \mathbf{r} is odd, the functions ψ_2 and ψ_1 must have different parity. Then the entire matrix element remains invariant under the replacement $\mathbf{r} \rightarrow (-\mathbf{r})$.

In deriving relations (106.1) the spin states of the electron were not taken into account, i.e. it was assumed that the spin state is not related to the orbital motion. In this case conditions (106.8) must be supplemented by the relation $\Delta s = 0$ which expresses spin conservation in the dipole transition. However, if the spin-orbit interaction cannot be disregarded, as, for example, in the case of heavy atoms and nuclei, then it is necessary to formulate selection rules for the total angular momentum J . Taking into account that in a dipole transition the quantum carries away an angular momentum equal to one, then, according to the rule of addition of angular momenta in quantum mechanics we obtain

$$\Delta j = 0, \pm 1 \quad (\text{excluding } 0 \rightarrow 0 \text{ transitions}). \quad (106.9)$$

In this case transitions with $\Delta j = 0$ are not forbidden, since the total angular momentum is not directly related to the parity of the state. The transition from the state $j_1 = 0$ into the state $j_2 = 0$ is forbidden, since in this case the total angular momentum conservation law cannot be satisfied.

In the case of magnetic dipole radiation the quantum also carried away an angular momentum equal to one. However, the magnetic dipole quantum has a parity opposite to the parity of the electric dipole quantum. This is associated with the fact that the magnetic moment operator does not change sign under the inversion of the system of coordinates, since the magnetic moment is a pseudovector. Consequently, the matrix elements of the magnetic moment operator are different from zero only for transitions between states of the same parity.

The electric quadrupole quantum carries away an angular momentum equal to two. In correspondence with this the total angular momentum selection rules are of the form

$$\Delta j = 0, \pm 1, \pm 2. \quad (106.10)$$

Transitions with the angular momenta

$$0 \rightarrow 0; \quad \frac{1}{2} \rightarrow \frac{1}{2}; \quad 0 \rightleftharpoons 1$$

are forbidden. The change in angular momentum in the emission given by relations (106.9) and (106.10) refers either to one particle, if only its state is changed, or to the system as a whole, for example to an atom or nucleus.

If the system is in a certain excited state and dipole transition to a lower energy state is forbidden, then the lifetime of the system in this excited state can be rather large. States of such a type are said to be metastable. In gases which are not very rarefied a metastable atom usually transfers its excitation energy in collisions with other atoms without emission.

Transitions associated with the angular momentum change $\Delta j \approx 4, 5$, which are strongly forbidden, are observed in nuclei. The lifetime of the nucleus with respect to such a transition for small excitation energies may reach several months. Such nuclei are said to be isomeric. They were first observed by Kurchatov and Rusinov.

§107. The photoelectric effect

The process of absorption of a photon by a bound particle when the energy of the photon exceeds the binding energy of the particle is called the photoelectric effect. In particular, in the photoelectric effect in an atom an electron in a state belonging to a discrete spectrum absorbs the photon and makes a transition to the continuous spectrum. The kinetic energy T of the electron knocked out of the atom is defined by the Einstein relation

$$T = \hbar\omega - I, \quad (107.1)$$

where I is the ionization energy of the atom.

The momentum excess arising when the photon is absorbed is transferred to the nucleus. The more strongly the electron is bound in the atom, the more easily the momentum is transformed to the nucleus. Hence it is to be expected that the probability of occurrence of the photoelectric effect will have a maximum value for the most strongly bound electrons, the electrons of the K-shell.

In what follows we shall restrict ourselves to the consideration of this case. The matrix element of the transition with the absorption of one quantum has the form (103.1). The wave functions ψ_1 and ψ_2 in the matrix element correspond respectively to the ground state of the electron in the atom and to a state belonging to the continuous spectrum. Since we do not take into account relativistic effects, it is obvious that the energy of the photon must in any case be small in comparison with the rest energy of the electron $\hbar\omega \ll mc^2$.

On the other hand, we exclude the region close to the threshold of the photoelectric effect, and assume that the energy of the photon is large in comparison with the ionization energy of the atom. Taking into account (107.1) and (38.17), these requirements lead to the inequality

$$T = \frac{p^2}{2m} \gg I = \frac{Z^2 e^4 m}{2\hbar^2} \quad \text{or} \quad \frac{Ze^2}{\hbar v} \ll 1. \quad (107.2)$$

According to the results of §84, the fulfillment of inequality (107.2) means that the Coulomb field acting on the electron can be considered as a small perturbation. Consequently, one can take a plane wave for the wave function ψ_2 of the free particle in the zero order approximation (disregarding the action of the Coulomb field on the free electron). The wave function of an electron in the K-shell can be written in the form of the hydrogen function with the effective nuclear charge Z (the action of other electrons on K-electrons is small). We then have

$$\psi_1 = \left(\frac{Z^3}{\pi a^3} \right)^{\frac{1}{2}} e^{-Zr/a}; \quad \psi_2 = \frac{1}{(2\pi\hbar)^{\frac{3}{2}}} e^{(i/\hbar)(\mathbf{p} \cdot \mathbf{r})}. \quad (107.3)$$

We normalize the wave function of the final state ψ_2 belonging to the continuous spectrum to the δ -function in momentum space. The transition probability per unit time, according to (56.8'), is equal to

$$dW = \frac{2\pi}{\hbar} |H'_{21}|^2 \delta(E_2 - E_1 - \hbar\omega) p^2 dp d\Omega. \quad (107.4)$$

Here $d\Omega$ is the element of solid angle characterizing the direction of the momentum \mathbf{p} of the emitted photoelectron, $E_1 = -I$, and $E_2 = p^2/2m$.

We integrate (107.4) over the energies of the final state

$$dW = \frac{2\pi m}{\hbar} |H'_{21}|^2 p d\Omega. \quad (107.5)$$

The value of the momentum \mathbf{p} is determined by relation (107.1). The integral in the matrix element (107.5) is easily calculated after the substitution of the wave functions given above. For this we write it in the form

$$H'_{21} = -\frac{e}{m(2\pi\hbar)} \left(\frac{Z^3}{\pi a^3 V \omega} \right)^{\frac{1}{2}} \int e^{(i/\hbar)(\mathbf{q} \cdot \mathbf{r})} \frac{\hbar}{i} (\mathbf{e} \cdot \nabla) e^{-Zr/a} dV,$$

where \mathbf{e} is the polarization vector of the photon.

Here we have introduced the vector $\mathbf{q} = \hbar \mathbf{k} - \mathbf{p}$ which represents the momentum transferred to the nucleus, and used the transverse property of electromagnetic waves $\mathbf{e} \cdot \mathbf{k} = 0$. Integrating by parts we obtain

$$H'_{21} = -\frac{e}{m(2\pi\hbar)} \left(\frac{Z^3}{\pi a^3 V \omega} \right)^{\frac{1}{2}} (\mathbf{e} \cdot \mathbf{p}) \int e^{-Zr/a} e^{(i/\hbar)(\mathbf{q} \cdot \mathbf{r})} dV.$$

Passing to spherical coordinates with the polar axis directed along the vector \mathbf{q} , we find

$$\begin{aligned} \int e^{-Zr/a} e^{(i/\hbar)(\mathbf{q} \cdot \mathbf{r})} dV &= \frac{2\pi\hbar}{iq} \int_0^\infty (e^{(i/\hbar)(\mathbf{q} \cdot \mathbf{r})} - e^{-(i/\hbar)(\mathbf{q} \cdot \mathbf{r})}) e^{-Zr/a} r dr = \\ &= \frac{8\pi a^3}{Z^3(1+q^2 a^2/Z^2 \hbar^2)^2}. \end{aligned}$$

Consequently, the matrix element H'_{21} is of the form

$$H'_{21} = -\frac{4e}{m\hbar} \left(\frac{a^3}{\pi Z^3 V \omega} \right)^{\frac{1}{2}} (\mathbf{e} \cdot \mathbf{p}) \frac{1}{(1+q^2 a^2/\hbar^2 Z^2)^2}. \quad (107.6)$$

We obtain the differential cross section for the photoelectric effect if we divide the transition probability per unit time (107.5) by the incident photon flux density. Since one quantum is absorbed and the process is normalized in such a way that in volume V there is one photon, then the incident photon flux density is equal to c/V . In correspondence with this we have

$$d\sigma = \frac{32 \times 137^4}{Z^3} \frac{p(\mathbf{p} \cdot \mathbf{e})^2 c^3}{(mc^2)^2 \hbar \omega} \left(\frac{e^2}{mc^2} \right)^2 \frac{d\Omega}{(1+q^2 a^2/\hbar^2 Z^2)^4}. \quad (107.7)$$

The constant $r_0 = e^2/mc^2 \sim 10^{-13}$ cm, as is well known, is called the classical radius of the electron. Expression (107.7) can be simplified. First of all, we denote the angle between the direction of momentum of the incident quantum and that of the emitted photoelectron by ϑ , i.e. the angle between the vectors \mathbf{k} and \mathbf{p} , and the angle between the (\mathbf{p}, \mathbf{k}) -plane and (\mathbf{e}, \mathbf{k}) -plane by φ . Then, writing $\hbar k = \kappa$, we have

$$\begin{aligned} \mathbf{p} \cdot \mathbf{e} &= p \sin \vartheta \cos \varphi, \\ q^2 &= p^2 + \kappa^2 - 2p\kappa \cos \vartheta. \end{aligned} \quad (107.8)$$

The expression $1 + q^2 a^2 / \hbar^2 Z^2$ involved in (107.7) can also be rewritten in a simpler form:

$$1 + \frac{q^2 a^2}{\hbar^2 Z^2} = \frac{a^2}{\hbar^2 Z^2} \left(\frac{Z^2 \hbar^2}{a^2} + q^2 \right) = \frac{a^2}{\hbar^2 Z^2} \left(\frac{Z^2 m^2 e^4}{\hbar^2} + p^2 + \kappa^2 - 2p\kappa \cos \vartheta \right).$$

Taking into account that $I = Z^2 e^4 m / 2\hbar^2$, it follows from relation (107.1) that

$$\frac{Z^2 m^2 e^4}{\hbar^2} + p^2 = 2m\hbar\omega = 2m\kappa c.$$

Then the preceding relation is rewritten as

$$1 + \frac{q^2 a^2}{\hbar^2 Z^2} = \frac{a^2}{\hbar^2 Z^2} \kappa (2mc + \kappa - 2p \cos \vartheta) = \frac{a^2}{\hbar^2 Z^2} 2m\hbar\omega (1 - \beta \cos \vartheta), \quad (107.9)$$

where $\beta = v/c$.

Here we have made use of the condition $\kappa c = \hbar\omega \ll mc^2$. The absolute value of the momentum \mathbf{p} involved in (107.5) can be replaced, according to (107.1) and (107.2), by the quantity $(2m\hbar\omega)^{1/2}$. Taking into account relations (107.9) and (107.8), we obtain for the differential cross section (107.7) the following expression:

$$d\sigma = 4\sqrt{2} \frac{Z^5}{137^4} r_0^2 \left(\frac{mc^2}{\hbar\omega} \right)^{\frac{7}{2}} \frac{\sin^2 \vartheta \cos^2 \varphi}{(1 - \beta \cos \vartheta)^4} d\Omega. \quad (107.10)$$

Since expression (107.10) is obtained in the non-relativistic approximation, it makes sense only to within terms of the first order with respect to β . Hence we have finally

$$d\sigma = 4\sqrt{2} \frac{Z^5}{137^4} r_0^2 \left(\frac{mc^2}{\hbar\omega} \right)^{\frac{7}{2}} \sin^2 \vartheta \cos^2 \varphi (1 + 4\beta \cos \vartheta) d\Omega. \quad (107.11)$$

From expression (107.11) it follows that photoelectrons are emitted mainly in the direction of polarization of the photon $\vartheta = \frac{1}{2}\pi$, $\varphi = 0$. Photoelectrons are not emitted in the direction of propagation of the quantum ($\vartheta=0$). As the energy of the quantum increases the maximum is considerably displaced in the forward direction. In order to obtain the total cross section for the photoelectric effect on the K-shell it is necessary to integrate (107.11) over all angles ϑ , φ and in addition to introduce the factor 2, since there are two electrons in the K-shell

$$\sigma = \frac{32\sqrt{2}}{3} \pi \frac{Z^5}{137^4} r_0^2 \left(\frac{mc^2}{\hbar\omega} \right)^{\frac{7}{2}}. \quad (107.12)$$

We see that the total cross section increases rapidly with increasing charge of the nucleus (as Z^5) and decreases with increasing frequency of the quantum (as $\omega^{-\frac{7}{2}}$).

§108. The scattering of light by atoms

As an important example of a process involving two photons we shall consider the quantum theory of the scattering of light by atoms. Let a photon with wave vector \mathbf{k}_1 be incident on an atom, and let a photon with wave vector \mathbf{k}_2 be emitted. We denote the corresponding frequencies by ω_1 and ω_2 , and the polarization vectors by \mathbf{e}_1 and \mathbf{e}_2 . If the frequency of the incident photon is equal to the frequency of the emitted photon, i.e. $\omega_1 = \omega_2$, then after scattering the atom returns to its initial state. Such scattering without change in the frequency is called coherent. We have seen in Part I that from the point of view of classical radiation theory only coherent scattering is possible. From the quantum-mechanical point of view scattering with a change in the frequency is as natural as coherent scattering, and was observed experimentally by Raman and independently by Mandelstam and Landsberg. It is called Raman scattering.

For generality we shall assume that the state of the atom changes in the act of scattering. We shall suppose that the energy of the incident photon is lower than the binding energy of the electron in the atom, which corresponds to the visible region of the spectrum. For photon energies large in comparison with this quantity the electron can be considered as free. However, we shall defer the consideration of the scattering of a photon by a free electron (the Compton effect) to §127.

If in the initial state the energy of the atom is E_1 and in the final state E_2 , then the energy conservation law gives

$$\hbar\omega_2 = \hbar\omega_1 + E_1 - E_2. \quad (108.1)$$

Let us write down the matrix elements of the interaction operator \hat{H}' (102.2) for the scattering process. Since the process involves two photons, it is necessary to take into account the contribution of the operator \hat{A}^2 to the matrix element. We denote the operator $-(e/mc)(\hat{\mathbf{p}} \cdot \hat{\mathbf{A}})$ by \hat{H}'_1 , and the operator $(e^2/2mc^2)\hat{A}^2$ by \hat{H}'_2 . Then the total perturbation operator is

$$\hat{H}' = \hat{H}'_1 + \hat{H}'_2. \quad (108.2)$$

Taking into account (102.5), it follows from (102.4) that the operator \hat{A}^2 gives a contribution to the process studied in the first approximation of perturbation theory

$$(\hat{H}'_2)_{21} = \frac{2\pi e^2}{mV} \frac{\hbar}{(\omega_1 \omega_2)^{\frac{1}{2}}} \int \psi_2^* e^{i(\mathbf{k}_1 - \mathbf{k}_2) \cdot \mathbf{r}} (\mathbf{e}_1 \cdot \mathbf{e}_2) \psi_1 dV. \quad (108.3)$$

For the frequencies considered the wavelength of the light is considerably larger than the size of the atom. In correspondence with this the exponential function in (108.3) can be set equal to unity. Also taking into account the orthogonality of the functions ψ_1 and ψ_2 , we have

$$(\hat{H}'_2)_{21} = \frac{2\pi e^2}{mV} \frac{\hbar}{(\omega_1 \omega_2)^{\frac{1}{2}}} (\mathbf{e}_1 \cdot \mathbf{e}_2) \delta_{12}. \quad (108.4)$$

The operator \hat{H}'_1 has matrix elements different from zero only for processes involving one photon. In the case of scattering the operator \hat{H}'_1 can give a contribution to the transition probability only in the second order approximation of perturbation theory. In §56 it was shown that in order to define the probability of a process in the second order approximation it is necessary to find the matrix elements of the operator \hat{H}'_1 corresponding to transitions into intermediate states.

Two types of intermediate states, over which the summation should be carried out, are possible in the process of scattering. (1) In the transition from the initial to the intermediate state of the first type the photon k_1 is absorbed and the atom makes a transition into a certain state which we shall characterize by the index i (energy E_i). In the subsequent transition from the intermediate into the final state the photon k_2 is emitted. According to (103.1) and (103.2) the matrix elements for the transition from the initial into the intermediate state and from the intermediate into the final state are of the form

$$\begin{aligned}
 (\hat{H}'_1)_{11} &= -\frac{e}{m} \left(\frac{2\pi\hbar}{V\omega_1} \right)^{\frac{1}{2}} \int \psi_i^* (\hat{\mathbf{p}} \cdot \mathbf{e}_1) e^{i\mathbf{k}_1 \cdot \mathbf{r}} \psi_1 dV, \\
 (\hat{H}'_1)_{21} &= -\frac{e}{m} \left(\frac{2\pi\hbar}{V\omega_2} \right)^{\frac{1}{2}} \int \psi_2^* (\hat{\mathbf{p}} \cdot \mathbf{e}_2) e^{-i\mathbf{k}_2 \cdot \mathbf{r}} \psi_i dV.
 \end{aligned}
 \tag{108.5}$$

(2) In the transition to the intermediate state of the second type the emission of the photon k_2 first occurs. Then the photon k_1 is absorbed and the atoms makes a transition from the intermediate into the final state. We recall that the energy conservation law holds only for the initial and final states. The matrix elements of the transition via the second intermediate state are written as

$$\begin{aligned}
 (\hat{H}'_1)_{11} &= -\frac{e}{m} \left(\frac{2\pi\hbar}{V\omega_2} \right)^{\frac{1}{2}} \int \psi_i^* (\hat{\mathbf{p}} \cdot \mathbf{e}_2) e^{-i\mathbf{k}_2 \cdot \mathbf{r}} \psi_1 dV, \\
 (\hat{H}'_1)_{21} &= -\frac{e}{m} \left(\frac{2\pi\hbar}{V\omega_1} \right)^{\frac{1}{2}} \int \psi_2^* (\hat{\mathbf{p}} \cdot \mathbf{e}_1) e^{i\mathbf{k}_1 \cdot \mathbf{r}} \psi_i dV.
 \end{aligned}
 \tag{108.6}$$

The constitutive matrix element Λ is given by formula (56.19)

$$\Lambda_{21} = \sum_i \left(\frac{(\hat{H}'_1)_{21}(\hat{H}'_1)_{11}}{E_{\text{init}} - E_i} + \frac{(\hat{H}'_1)_{21}(\hat{H}'_1)_{11}}{E_{\text{init}} - (E_i + \hbar\omega_1 + \hbar\omega_2)} \right). \tag{108.7}$$

The energy of the initial state is made up of the energy of the atom E_1 and the energy of the incident quantum $\hbar\omega_1$, i.e.

$$E_{\text{init}} = E_1 + \hbar\omega_1. \tag{108.8}$$

The energy of the intermediate state of the second type involves, in addition to the energy of the atom, the total energy of the two photons.

Substituting expressions (108.5) and (108.6) into (108.7) we obtain

$$\Lambda_{21} = \frac{e^2 2\pi\hbar}{m^2 V (\omega_1 \omega_2)^{\frac{1}{2}}} \sum_i \left(\frac{(\mathbf{e}_2 \cdot \hat{\mathbf{p}})_{2i} (\mathbf{e}_1 \cdot \hat{\mathbf{p}})_{i1}}{E_1 - E_i + \hbar\omega_1} + \frac{(\mathbf{e}_1 \cdot \hat{\mathbf{p}})_{2i} (\mathbf{e}_2 \cdot \hat{\mathbf{p}})_{i1}}{E_1 - E_i - \hbar\omega_2} \right). \tag{108.9}$$

Here we have substituted unity for the exponential expressions $e^{i\mathbf{k}_1 \cdot \mathbf{r}}$ and $e^{-i\mathbf{k}_2 \cdot \mathbf{r}}$. The summation over the energy states of the atom must also involve integration over states belonging to the continuous spectrum. The total matrix element for the process considered is obtained by adding to (108.9) the matrix element (108.4)

$$M_{21} = \frac{2\pi e^2}{mV} \frac{n}{(\omega_1 \omega_2)^{\frac{1}{2}}} \left[\frac{1}{m} \sum_i \left(\frac{(\mathbf{e}_2 \cdot \hat{\mathbf{p}})_{2i} (\mathbf{e}_1 \cdot \hat{\mathbf{p}})_{i1}}{E_1 - E_i + \hbar \omega_1} + \right. \right. \\ \left. \left. + \frac{(\mathbf{e}_1 \cdot \hat{\mathbf{p}})_{2i} (\mathbf{e}_2 \cdot \hat{\mathbf{p}})_{i1}}{E_1 - E_i - \hbar \omega_2} \right) + \delta_{12} (\mathbf{e}_1 \cdot \mathbf{e}_2) \right]. \quad (108.10)$$

The transition probability per unit time is given, as always, by the formula

$$dW = \frac{2\pi}{\hbar} |M_{21}|^2 \rho(\omega_2) d\Omega. \quad (108.11)$$

Here $\rho(\omega_2)$, as in §103, denotes the number of field oscillators in volume V corresponding to unit energy interval (see (103.4)). The solid angle element $d\Omega$ characterizes the direction of the momentum of the scattered photon.

Dividing the transition probability per unit time (108.11) by the incident photon flux density which, as in the preceding section, is equal to c/V , we obtain the expression for the differential cross section for the process

$$d\sigma = r_0^2 \frac{\omega_2}{\omega_1} \left| \frac{1}{m} \sum_i \left(\frac{(\mathbf{e}_2 \cdot \hat{\mathbf{p}})_{2i} (\mathbf{e}_1 \cdot \hat{\mathbf{p}})_{i1}}{E_1 - E_i + \hbar \omega_1} + \frac{(\mathbf{e}_1 \cdot \hat{\mathbf{p}})_{2i} (\mathbf{e}_2 \cdot \hat{\mathbf{p}})_{i1}}{E_1 - E_i - \hbar \omega_2} + \delta_{12} (\mathbf{e}_1 \cdot \mathbf{e}_2) \right) \right|^2 d\Omega. \quad (108.12)$$

It is clear that formula (108.12) is inapplicable in the case of perfect resonance $\hbar \omega_1 = E_1 - E_i$.

Formula (108.12) characterizes the scattering capacity of the atom as a function of the frequency of the incident light, and hence is called, as in classical electrodynamics, a dispersion formula. The last term of (108.12) is different from zero only for coherent scattering $\omega_1 = \omega_2$, for which the initial and final states of the atom are the same. If the initial and final states of the atom are not identical, then the frequency of the scattered radiation is shifted with respect to the frequency of the incident radiation by an amount corresponding to the difference between the energy states of the atom (108.1). Scattering of such a type is called the Raman effect.

Formula (108.12) can be rewritten in a somewhat different form if the matrix elements of the momentum are expressed in terms of the matrix elements of the coordinates of the electron. First of all, it is convenient to write the last term in the bracket in (108.12) in the same form as the two preceding terms.

For this we note* that, as a result of the commutation of the correspond-

* See the monograph of A.I. Akhiezer and V.B. Berestetskii, *Quantum electrodynamics* (Interscience Publishers, New York, 1965).

ing components of the operators $\hat{\mathbf{p}}$ and $\hat{\mathbf{r}}$, the scalar product $\mathbf{e}_1 \cdot \mathbf{e}_2$ can be expressed as

$$(\mathbf{e}_1 \cdot \mathbf{e}_2) = \frac{i}{\hbar} [(\mathbf{e}_1 \cdot \hat{\mathbf{p}})(\mathbf{e}_2 \cdot \mathbf{r}) - (\mathbf{e}_2 \cdot \mathbf{r})(\mathbf{e}_1 \cdot \hat{\mathbf{p}})] . \quad (108.13)$$

We shall take into account the fact that the scalar product $\mathbf{e}_1 \cdot \mathbf{e}_2$ is involved in (108.12) only for coherent scattering if we take from (108.13) the matrix element of the transition from the initial into the final state. This matrix element is, by virtue of the orthogonality of the corresponding wave functions, different from zero only under the condition that the initial state of the atom is the same as the final state

$$\begin{aligned} \delta_{21}(\mathbf{e}_1 \cdot \mathbf{e}_2) &= \frac{i}{\hbar} [(\mathbf{e}_1 \cdot \hat{\mathbf{p}})(\mathbf{e}_2 \cdot \mathbf{r}) - (\mathbf{e}_2 \cdot \mathbf{r})(\mathbf{e}_1 \cdot \hat{\mathbf{p}})]_{21} = \\ &= \frac{i}{\hbar} \sum_i ((\mathbf{e}_1 \cdot \hat{\mathbf{p}})_{2i}(\mathbf{e}_2 \cdot \mathbf{r})_{i1} - (\mathbf{e}_2 \cdot \mathbf{r})_{2i}(\mathbf{e}_1 \cdot \hat{\mathbf{p}})_{i1}) . \end{aligned} \quad (108.14)$$

We substitute the expression obtained into (108.12). The matrix elements of the momentum, according to (49.5), are expressed in terms of the matrix elements of the coordinate

$$(\mathbf{p})_{21} = \frac{i}{\hbar} m(E_2 - E_1)(\mathbf{r})_{21} . \quad (108.15)$$

Adding up in (108.12) the corresponding matrix elements in brackets and taking into account the energy conservation law (108.1), we obtain the following expression for the bracket involved in (108.12):

$$[\dots] = \frac{m}{\hbar} \sum_i \left(\frac{\omega_2(E_i - E_1)(\mathbf{e}_2 \cdot \mathbf{r})_{2i}(\mathbf{e}_1 \cdot \mathbf{r})_{i1}}{E_1 - E_i + \hbar\omega_1} + \frac{\omega_2(E_2 - E_i)(\mathbf{e}_1 \cdot \mathbf{r})_{2i}(\mathbf{e}_2 \cdot \mathbf{r})_{i1}}{E_1 - E_i - \hbar\omega_2} \right) . \quad (108.16)$$

This expression can be simplified if the following term, equal to zero, is added to it:

$$\begin{aligned} \frac{m\omega_2}{\hbar} [(\mathbf{e}_1 \cdot \mathbf{r})(\mathbf{e}_2 \cdot \mathbf{r}) - (\mathbf{e}_2 \cdot \mathbf{r})(\mathbf{e}_1 \cdot \mathbf{r})]_{21} &= \\ &= \frac{m\omega_2}{\hbar} \sum_i ((\mathbf{e}_2 \cdot \mathbf{r})_{2i}(\mathbf{e}_1 \cdot \mathbf{r})_{i1} - (\mathbf{e}_1 \cdot \mathbf{r})_{2i}(\mathbf{e}_2 \cdot \mathbf{r})_{i1}) . \end{aligned}$$

We then obtain

$$[\dots] = m\omega_1\omega_2 \sum_i \left(\frac{(\mathbf{e}_2 \cdot \mathbf{r})_{2i}(\mathbf{e}_1 \cdot \mathbf{r})_{i1}}{E_1 - E_i + \hbar\omega_1} + \frac{(\mathbf{e}_1 \cdot \mathbf{r})_{2i}(\mathbf{e}_2 \cdot \mathbf{r})_{i1}}{E_1 - E_i - \hbar\omega_2} \right). \quad (108.17)$$

Substituting (108.17) into (108.12) we find the final expression for the differential scattering cross section

$$d\sigma = \frac{e^4}{c^4} \omega_1 \omega_2^3 \left| \sum_i \left(\frac{(\mathbf{e}_2 \cdot \mathbf{r})_{2i}(\mathbf{e}_1 \cdot \mathbf{r})_{i1}}{E_1 - E_i + \hbar\omega_1} + \frac{(\mathbf{e}_1 \cdot \mathbf{r})_{2i}(\mathbf{e}_2 \cdot \mathbf{r})_{i1}}{E_1 - E_i - \hbar\omega_2} \right) \right|^2 d\Omega. \quad (108.18)$$

A remarkable feature of the formulae obtained is the fact that for coherent scattering they are the same as the classical formulae in §36 of Part I. Formula (108.18) is widely used in practice, since the study of Raman scattering turned out to be a very effective method of investigating the energy levels and other properties of complex molecules.

We note also that the matrix elements of the atomic dipole moment induced by light can be expressed on the basis of the relations obtained. In turn, as was shown in §34 of Part IV, it is easy to find the relation between the dielectric constant and the induced dipole moment for a rarefied gas. Hence formula (108.18) is the basis for the quantum-mechanical calculation of the dielectric constant ϵ and correspondingly of the refractive index $n = \epsilon^{1/2}$. In particular, for coherent scattering the following expression is obtained for the quantity n^2 :

$$n^2 = 1 + \frac{4\pi e^2 N}{m} \sum_i \frac{f_i}{\omega_{i1}^2 - \omega^2}, \quad (108.19)$$

where the quantity $f_i = 2m\hbar^{-1}\omega_{i1}|x_{i1}|^2$ is called the oscillator strength, $\hbar\omega_{i1} = E_i - E_1$, and N is the number of atoms per unit volume. We have chosen the direction of polarization of the photons as the x -axis. The following relation holds for the oscillator strengths f_i .*

$$\sum_i f_i = 1.$$

As was pointed out earlier the expressions obtained make no sense near a

* See, for example, H. Bethe and E. Salpeter, *Quantum mechanics of one and two electron systems*, Handbuch der Physik, Bd. 35 (Springer-Verlag, Berlin, 1957).

resonance. Investigation of this phenomenon, called resonance fluorescence, is impossible without introducing the concept of linewidth, to which the next section is devoted. There the formulae of scattering theory taking into account the linewidth will be presented.

§109. The theory of the natural linewidth

In §103–105 we have found the total probability of emission of a photon by an atomic system. A more precise investigation of the corresponding equations makes it possible to find also the frequency distribution of the intensity of the radiation, i.e. to determine the form of the spectral line of the radiation. We recall that such a problem can also be solved within the framework of classical concepts (see Part I). In this case in order to obtain the natural form of a line it was necessary to take into account the damping of the amplitude of the radiation oscillator.

Although the interaction of charged particles with the electromagnetic field is weak, we cannot confine ourselves to the usual approximation of perturbation theory in considering the form of a spectral line. As a matter of fact, finding the form of a line is associated with the necessity of taking into account the decay of the initial state of the atomic system. Such a decay takes place over a rather long time t and, naturally, cannot be taken into account by the methods of perturbation theory (see §56).

We shall proceed from the general system of eqs. (55.5) for the amplitudes c_m of unperturbed states (atom and radiation field). We shall denote the amplitude of the initial state of the system by $\varphi'(t)$. In this state there are no photons, and the atom is in an excited state with energy E_2 . In the system of eqs. (55.5) it is sufficient to take into account only those states whose energy is approximately the same as the energy of the initial state. Only these states play an important role. We shall assume for simplicity that the state with energy E_2 is the first excited state of the atom, and that the energy of the ground state is equal to E_1

$$E_2 - E_1 = \hbar\omega_0.$$

The emitted photon has a frequency ω close to the frequency ω_0 . We shall denote the amplitude of the state arising in the transition $2 \rightarrow 1$ by $f'(t)$ (the atom makes a transition into the ground state and emits a photon with wave vector \mathbf{k} , frequency ω_λ and polarization \mathbf{e}).

In the case given the system of eqs. (55.5) will have the form

$$\begin{aligned}
 i\hbar \dot{f}'_{\lambda}(t) &= \langle 1, 1_{\lambda} | H' | 2, 0 \rangle e^{-i(\omega_{\lambda} - \omega_0)t} \varphi'(t), \\
 i\hbar \dot{\varphi}'(t) &= \sum_{\lambda} \langle 2, 0 | H' | 1, 1_{\lambda} \rangle e^{i(\omega_{\lambda} - \omega_0)t} f'_{\lambda}(t).
 \end{aligned}
 \tag{109.1}$$

From the initial conditions it follows that $f'_{\lambda}(0) = 0$, $\varphi'(0) = 1$. We introduce the notation

$$\begin{aligned}
 f_{\lambda}(t) &= f'_{\lambda}(t) e^{-i\omega_{\lambda}t}, & \langle 1, 1_{\lambda} | H' | 2, 0 \rangle &= H'_{12}, \\
 \varphi(t) &= \varphi'(t) e^{-i\omega_0t}, & \langle 2, 0 | H' | 1, 1_{\lambda} \rangle &= H'_{21} = H'^*_{12}.
 \end{aligned}
 \tag{109.2}$$

The Hamiltonian of the interaction of charged particles with the electromagnetic field, \hat{H}' , is given by expression (102.2). The form of the line is determined by the value of $|f'_{\lambda}(t)|^2 = |f_{\lambda}(t)|^2$ for $t \rightarrow \infty$.

Substituting expression (109.2) into (109.1) we obtain the system of equations of the amplitudes $f_{\lambda}(t)$ and $\varphi(t)$

$$\begin{aligned}
 i\hbar \dot{f}_{\lambda}(t) &= H'_{12} \varphi(t) + \hbar \omega_{\lambda} f_{\lambda}(t), \\
 i\hbar \dot{\varphi}(t) &= \hbar \omega_0 \varphi(t) + \sum_{\lambda} H'^*_{12} f_{\lambda}(t).
 \end{aligned}
 \tag{109.3}$$

The system of eqs. (109.3) is conveniently solved by means of a Laplace transformation. We denote the amplitudes in the Laplace representation by $f_{\lambda}(p)$

$$f_{\lambda}(p) = \int_0^{\infty} f_{\lambda}(t) e^{-pt} dt,
 \tag{109.4}$$

$$f_{\lambda}(t) = \frac{1}{2\pi i} \int_{-i\infty+\delta}^{i\infty+\delta} f_{\lambda}(p) e^{pt} dp.$$

Taking into account the initial conditions, we have

$$ipf_{\lambda}(p) = \frac{1}{\hbar} H'_{12} \varphi(p) + \omega_{\lambda} f_{\lambda}(p), \quad (109.5)$$

$$ip\varphi(p) = i + \omega_0 \varphi(p) + \frac{1}{\hbar} \sum_{\lambda} H'^*_{12} f_{\lambda}(p).$$

Eliminating the amplitude $\varphi(p)$ from this system, we obtain the equation for the function $f_{\lambda}(p)$

$$(ip - \omega_{\lambda})(ip - \omega_0)f_{\lambda}(p) = \frac{i}{\hbar} H'_{12} + \frac{H'_{12}}{\hbar^2} \sum_{\lambda} H'^*_{12} f_{\lambda}(p). \quad (109.6)$$

Multiplying the left-hand and right-hand sides of the equation by the function H'^*_{12} and summing over λ , we find the expression for the sum

$$\frac{1}{\hbar} \sum_{\lambda} H'^*_{12} f_{\lambda}(p) = \frac{\gamma}{2(ip - \omega_0 + \frac{1}{2}i\gamma)}, \quad (109.7)$$

where γ is defined by the relation

$$\frac{1}{\hbar^2} \sum_{\lambda} \frac{|H'_{12}|^2}{ip - \omega_{\lambda}} = -\frac{1}{2}i\gamma. \quad (109.8)$$

We pass from a discrete to a continuous distribution of frequencies and replace the summation by integration. The denominator of the left-hand side of expression (109.8) contains an infinitely small imaginary positive addition. The integration over the frequency ω amounts to taking the integral in the sense of the principal value and the semiresidue at the point $\omega_{\lambda} = ip$

$$\int \frac{F(\omega) d\omega}{ip - \omega} = P \int \frac{F(\omega) d\omega}{ip - \omega} - i\pi F(ip), \quad (109.9)$$

where F is an arbitrary function.

We disregard the integral taken in the sense of the principal value (it gives a small real shift of the spectrum of radiation frequencies). Since in integrating over the Laplace variable p the basic contribution is given by the region where $ip \approx \omega_0$, in finding the value of γ we can immediately take the residue at this point.

We have finally

$$\gamma = \frac{2\pi}{\hbar^2} \sum \int d\Omega |H'_{12}|^2 \frac{\omega_0^2 V}{(2\pi c)^3}. \quad (109.10)$$

On the right-hand side the summation is carried out over the polarizations and the integration over the directions of the vector \mathbf{k} of the emitted photon. From comparison of (109.10) with the expressions obtained in § 103 it follows that the quantity γ is the total probability of emission of a photon by the atom per unit time.

Substituting expression (109.7) into (109.6) we find the amplitude $f_\lambda(p)$

$$f_\lambda(p) = \frac{iH'_{12}}{\hbar(ip - \omega_0 + \frac{1}{2}i\gamma)(ip - \omega_\lambda)}. \quad (109.11)$$

From the transform (109.11) we find the original, the function $f_\lambda(t)$ (see (109.4)). Closing the integration contour in the left half-plane of the complex variable p and determining the residues, we find

$$f_\lambda(t) = \frac{H'_{12} e^{-i\omega_\lambda t}}{(\omega_\lambda - \omega_0 + \frac{1}{2}i\gamma)} [e^{i(\omega_\lambda - \omega_0)t - \frac{1}{2}\gamma t} - 1]. \quad (109.12)$$

The expression $|f_\lambda(t)|^2$, taken after the lapse of a sufficiently large time $t \gg \gamma^{-1}$, defines the probability of emission by the atom of a photon with given polarization and given wave vector \mathbf{k} , i.e. defines the form of the emission line

$$|f_\lambda(\infty)|^2 = \frac{|H'_{12}|^2}{\hbar^2} \frac{1}{(\omega_\lambda - \omega_0)^2 + \frac{1}{4}\gamma^2}. \quad (109.13)$$

The intensity of emission by the atom of a photon with given frequency ω , $J(\omega)$, is obtained from (109.13) multiplying $|f_\lambda(\infty)|^2$ by $\hbar\omega\rho(\omega)$ (see (103.4)) and summing and integrating over the polarizations and the directions of the vector \mathbf{k} of the emitted photon. Taking into account (109.10) we have

$$J(\omega)d\omega = \frac{\gamma}{2\pi} \frac{\hbar\omega d\omega}{(\omega - \omega_0)^2 + \frac{1}{4}\gamma^2}. \quad (109.14)$$

The intensity distribution of the radiation is of a dispersion character, and the distribution width γ is equal to the total probability of emission per unit time. The relation obtained between the distribution width and the transition probability is in correspondence with the general Heisenberg uncertainty relation for time and energy $\Delta E \Delta t > \hbar$ (see § 34). Here ΔE is the uncertainty in the energy of the excited state, $\Delta E \sim \hbar\Delta\omega$, $\Delta\omega$ is the distribution width

and Δt is the mean lifetime of the atom in the excited state; $\Delta t \sim \gamma^{-1}$, since the emission proceeds in a time of the order of γ^{-1} . Hence it follows that $\Delta\omega \sim \gamma$ in correspondence with the result obtained.

In the case where the transition takes place not between the first excited and ground levels but between arbitrary i th and k th levels, the width of the transition line γ_{ik} is equal to the sum of the widths γ_i and γ_k of the levels

$$\gamma_{ik} = \gamma_i + \gamma_k. \quad (109.15)$$

Each of the level widths γ_i and γ_k is equal to the sum of the probabilities of transition from the given level to all lower levels.

The width considered here is called 'natural', since it is determined by the process of emission itself, by the radiative reaction. In addition there are other mechanisms of spectral-line broadening, which usually lead to more noticeable effects. Thus, for example, in a gaseous system the collision broadening and the Doppler broadening are considerable*. The collision broadening is due to collisions between the molecules. Indeed, collisions interrupt the process of emission. Hence, if τ is the lifetime of the atom with respect to collisions (the mean time between collisions), then, as follows from the uncertainty relation for time and energy, the linewidth is of the order of τ^{-1} .

By taking into account the linewidth the results obtained in the preceding section can be extended to the region near resonance, where $\hbar\omega \approx E_n - E_1$, and where E_n is the energy of one of the atomic levels. In this case the basic contribution in the summation is given only by states with the 'resonance' energy E_n . Thus for the differential coherent scattering cross section we have, instead of (108.18),

$$d\sigma = \left(\frac{e\omega}{c}\right)^4 \frac{\sum_i (\mathbf{e}_2 \cdot \mathbf{r})_{2i} (\mathbf{e}_1 \cdot \mathbf{r})_{i1}}{(E_1 - E_n + \hbar\omega)^2 + \frac{1}{4}\Gamma_n^2} d\Omega, \quad (109.16)$$

where Γ_n is the total width of the n th level. The summation is carried out over all states with energy E_n .

Carrying out the summation and averaging over the initial states and summing over the final states of the system, we obtain

$$\sigma = \pi\lambda^2 \frac{2j_n + 1}{2j_1 + 1} \frac{\gamma_n^2}{(E_1 - E_n + \hbar\omega)^2 + \frac{1}{4}\Gamma_n^2}, \quad (109.17)$$

* This problem is considered in detail, for example, in the book of I.I. Sobelman, *Vvedenie v teoriyu atomnykh spektrov (Introduction to the theory of atomic spectra)* (Nauka, Moscow, 1963).

where γ_n is the natural width of the n th level, $\lambda = c/\omega$, and j_1 and j_n are the total angular momenta of the initial and the n th atomic states. The cross section reaches a maximum value equal to $4\pi\lambda^2(2j_n+1)/(2j_1+1)$ for perfect resonance and $\gamma_n = \Gamma_n$.

Relativistic Quantum Mechanics

§110. The relativistic wave equation for a particle of zero spin

So far we have restricted ourselves to the study of the properties of particles moving with velocities small in comparison with the velocity of light.

Indeed, in obtaining the Schrödinger equation we have written the non-relativistic Hamiltonian of a particle in an external potential field

$$H = \frac{\mathbf{p}^2}{2m} + U(\mathbf{r})$$

and have replaced the corresponding quantities in it by operators. To obtain the relativistic theory, use should be made of the same scheme as that developed in §27. Namely, to construct the wave equation one has to use the relativistic expression for the Hamiltonian. For generality we shall immediately assume that the particle moves in an external electromagnetic field. Then its Hamiltonian has the form of (23.17) of Part II. Carrying out the replacement of corresponding quantities by operators, i.e. $H \rightarrow i\hbar(\partial/\partial t)$, $\mathbf{p} \rightarrow -i\hbar\nabla$, we obtain the equation

$$\left(i\hbar \frac{\partial}{\partial t} - e\varphi\right)^2 \psi = c^2 \left(-i\hbar\nabla - \frac{e}{c}\mathbf{A}\right)^2 \psi + m^2 c^4 \psi. \quad (110.1)$$

Equation (110.1) is called the Klein–Gordon–Fock equation. The relativistic

invariance of this equation is evident. The Klein—Gordon—Fock equation is a second order wave equation.

Since the relativistic Hamiltonian goes over in the limit into that of classical mechanics, it is natural to assume that for $c \rightarrow \infty$ the Klein—Gordon—Fock equation will go over into the Schrödinger equation. We shall show this.

Since the zero point energies in non-relativistic theory and the theory of relativity differ by mc^2 , it is convenient to introduce a transformation of the wave function ψ by means of the following relation:

$$\psi(x, t) = \psi'(x, t) e^{-imc^2 t/\hbar}.$$

Substituting into (110.1) and calculating the derivatives with respect to time, we obtain

$$2i\hbar mc^2 \frac{\partial \psi'}{\partial t} - \hbar^2 \frac{\partial^2 \psi'}{\partial t^2} - 2e\varphi \left[mc^2 \psi' + i\hbar \frac{\partial \psi'}{\partial t} \right] + e^2 \varphi^2 \psi' = c^2 \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right)^2 \psi'. \quad (110.2)$$

We retain only terms proportional to c^2 in this equation. Upon dividing both sides of the equation by $2mc^2$ we arrive at the ordinary Schrödinger equation

$$i\hbar \frac{\partial \psi'}{\partial t} = \frac{\left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right)^2}{2m} \psi' + e\varphi \psi'. \quad (110.3)$$

Thus we have shown that the Klein—Gordon—Fock equation goes over into the Schrödinger equation in the non-relativistic limit.

Equation (110.1), like the Schrödinger equation, defines the development of a process in time. The state of the particle is, as before, characterized by the wave function $\psi(x, y, z, t)$. This function depends on the coordinates x, y, z, t and contains no spin variables. Hence it is clear that the Klein—Gordon—Fock equation defines the behaviour of a spin zero particle. In order that this equation may describe particles of spin different from zero it must somehow be modified.

Since the Klein—Gordon—Fock equation is relativistically invariant, the wave function can be multiplied only by a certain constant phase factor in Lorentz transformations. From normalization considerations it follows that this factor must be equal to +1 (but not -1, since the Lorentz transformation is continuous). Under space reflection of the coordinates the wave function ψ can be multiplied by +1 or -1. In other words, under the action of the parity operator the wave function can transform in two ways:

$$\hat{I}\psi(x, y, z, t) = +\psi(-x, -y, -z, t),$$

$$\hat{I}\psi(x, y, z, t) = -\psi(-x, -y, -z, t).$$

Thus the wave function ψ can be either a scalar or a pseudoscalar. For this reason the Klein–Gordon–Fock equation is often said to be a scalar equation.

As an example of the integration of eq. (110.1) we shall consider the case of a free particle. Then the Klein–Gordon–Fock equation can be written in the form

$$-\hbar^2 \frac{\partial^2 \psi}{\partial t^2} = -c^2 \hbar^2 \nabla^2 \psi + m^2 c^4 \psi. \quad (110.4)$$

We see the solution of eq. (110.4) in the form

$$\psi = e^{-iEt/\hbar} \psi_1(x, y, z).$$

Then for the function ψ_1 we find

$$E^2 \psi_1 = -c^2 \hbar^2 \nabla^2 \psi_1 + m^2 c^4 \psi_1.$$

Rewriting this equation as follows

$$\nabla^2 \psi_1 + \frac{E^2 - m^2 c^4}{c^2 \hbar^2} \psi_1 = 0, \quad (110.5)$$

we easily find that its solution is a plane wave of the form

$$\psi_1 = a e^{i(\mathbf{p} \cdot \mathbf{r})/\hbar}.$$

Substituting this value for ψ_1 into (110.5), we arrive at the relativistic relation between the energy E and momentum p of a free particle

$$E^2 = c^2 p^2 + m^2 c^4.$$

This relation is the same as the usual formula for energy in the theory of relativity.

§111. The charge density and probability current for particles of zero spin

We now turn to finding the charge density and probability current for particles described by the scalar wave equation (110.1). The derivation of expressions for these quantities is carried out according to the same scheme

as for the Schrödinger equation. Namely, we multiply the Klein-Gordon-Fock equation

$$\left(i\hbar \frac{\partial}{\partial t} - e\varphi\right)^2 \psi - c^2 \left(-i\hbar \nabla - \frac{e}{c} \mathbf{A}\right)^2 \psi - m^2 c^4 \psi = 0 \quad (111.1)$$

by the adjoint wave function ψ^* .

We multiply the equation conjugate to eq. (111.1) by the wave function ψ . We subtract the second equation from the first. As a result we find

$$-\hbar^2 \left[\psi^* \frac{\partial^2 \psi}{\partial t^2} - \psi \frac{\partial^2 \psi^*}{\partial t^2} \right] - 2i\hbar e\varphi \left(\psi^* \frac{\partial}{\partial t} \psi + \psi \frac{\partial}{\partial t} \psi^* \right) - \quad (111.2)$$

$$-\hbar^2 c^2 [\psi \nabla^2 \psi^* - \psi^* \nabla^2 \psi] - i\hbar e c [\psi^* (\nabla \mathbf{A} + \mathbf{A} \nabla) \psi + \psi (\nabla \mathbf{A} + \mathbf{A} \nabla) \psi^*] = 0.$$

The first expression in the bracket is easily transformed into the form

$$\psi^* \frac{\partial^2 \psi}{\partial t^2} - \psi \frac{\partial^2 \psi^*}{\partial t^2} = \frac{\partial}{\partial t} \left[\psi^* \frac{\partial \psi}{\partial t} - \psi \frac{\partial}{\partial t} \psi^* \right].$$

The second bracket is simply the derivative with respect to time of the product $\psi^* \psi$. By means of the vector relation

$$\psi \nabla^2 \psi^* = \psi \nabla \cdot \nabla \psi^* = \nabla \cdot (\psi \nabla \psi^*) - \nabla \psi \nabla \psi^*$$

we find that

$$\psi \nabla^2 \psi^* - \psi^* \nabla^2 \psi = \nabla \cdot [\psi \nabla \psi^* - \psi^* \nabla \psi].$$

Finally, the last expression in the bracket of formula (111.2) is brought by means of the relation

$$\psi^* (\nabla \mathbf{A}) \psi = \psi^* \nabla \cdot (\psi \mathbf{A}) = \nabla \cdot (\psi^* \psi \mathbf{A}) - \psi \mathbf{A} \nabla \psi^*$$

to the form

$$\psi^* (\nabla \mathbf{A} + \mathbf{A} \nabla) \psi + \psi (\nabla \mathbf{A} + \mathbf{A} \nabla) \psi^* = 2 \nabla \cdot (\psi \psi^* \mathbf{A}).$$

If one multiplies all terms of eq. (111.2) by the quantity $e/2i\hbar mc^2$ and makes use of the transformations shown, then eq. (111.2) can be written in the form

$$\nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t} = 0,$$

where the charge density ρ is equal to

$$\rho = \frac{e\hbar}{2imc^2} \left[\psi \frac{\partial \psi^*}{\partial t} - \psi^* \frac{\partial \psi}{\partial t} \right] - \frac{e^2}{mc^2} \psi \psi^* \varphi. \quad (111.3)$$

Then we have for the current density the expression

$$\mathbf{j} = \frac{i\hbar e}{2m} [\psi \nabla \psi^* - \psi^* \nabla \psi] - \frac{e^2}{mc^2} \psi \psi^* \mathbf{A}. \quad (111.4)$$

Let us dwell on the meaning of the results obtained. In non-relativistic theory the charge density ρ can be written in the form

$$\rho(x, y, z, t) = eW(x, y, z, t),$$

where $W(x, y, z, t)$ is the probability current, which is in essence a positive quantity. Clearly, relation (111.3) cannot be interpreted in such a way. The expression for ρ can be made negative by a proper choice of the function ψ at the initial instant of time.

Indeed, since the Klein—Gordon—Fock equation is a second order equation with respect to time, arbitrary values of the ψ -function itself and of its derivative with respect to time can be given at the initial instant of time. Choosing different ψ and $\partial\psi/\partial t$ it is possible to arrive at positive as well as negative values of the quantity ρ .

We shall now show that the quantity ρ/e has indeed as its non-relativistic limit the product $\psi^*\psi$. Let the following relation be fulfilled for the derivative of the wave function ψ with respect to time:

$$i\hbar \frac{\partial \psi}{\partial t} = E\psi.$$

In this case the expression for the charge density ρ can be written in the form

$$\rho = \frac{eE}{mc^2} \psi^* \psi - \frac{e^2 \varphi}{mc^2} \psi^* \psi.$$

If we separate the rest energy from the energy E , i.e. if we set $E = mc^2 + E'$, then in this case we easily obtain

$$\frac{\rho}{e} = \psi^* \psi \left[1 + \frac{E' - e\varphi}{mc^2} \right].$$

If the quantity $E' - e\varphi \ll mc^2$, then we have the correct expression for the non-relativistic limit of the quantity ρ/e . We see that in the case of the Klein—Gordon—Fock equation one cannot introduce a positively defined probability

current. This fact was the reason why, for a long time, the Klein–Gordon–Fock equation was not applied to real objects.

§112. The concept of the nuclear force field

The Klein–Gordon–Fock equation was subsequently given a new, completely different physical interpretation.

We already know that in addition to electrical interactions other forms of interaction occur in nature. In particular, one such interaction, which does not depend on the electric charge e , is the strong nuclear interaction. It seemed natural to assume that the nuclear interaction could be associated with the presence of a special nucleon charge g in nucleons. One can then try to describe the nuclear interaction by analogy with the interaction of electric charges, introducing the concept of a nuclear force field. This field should be described by a potential similar to the potential φ of the electric field. The attempt was made to give up the interpretation of the Klein–Gordon–Fock equation as the equation for the wave function of one particle. Instead it was proposed to consider the function ψ as the potential of the nuclear field produced by nucleons. Just as photons are quantum particles corresponding to the electromagnetic field, there correspond to the nuclear field π -mesons.

In §67 we have already dwelt on the obvious interpretation of the π -meson exchange and photon exchange as the sources, respectively, of the strong nucleon interaction and the electromagnetic interaction.

Carrying this analogy further, we can go on to write the equation for the nuclear field potential. As such an equation we take the Klein–Gordon–Fock equation in the form

$$\nabla^2 \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} - \kappa^2 \psi = 0, \quad \kappa^2 = m^2 c^2 / \hbar^2, \quad (112.1)$$

where m is the mass of the π -meson.

We shall not consider the quantum theory of nuclear forces, in which mesons are the elementary excitations of a certain field, like photons in the quantum theory of the electromagnetic field.

Since we are interested only in the qualitative aspect of the subject, we shall carry out our reasoning by analogy with the classical theory of the electrostatic field. Assuming that the nuclear field does not depend on time, we write for its potential ψ the equation

$$\nabla^2 \psi - \kappa^2 \psi = 0. \quad (112.2)$$

This equation is a certain analogue of the equation of the electrostatic field and goes over into the latter for $m \rightarrow 0$ (see below). As is well known, in the presence of point charges the equation of the electrostatic field is of the form

$$\nabla^2 \varphi = -4\pi e \delta(\mathbf{r}).$$

Hence in the presence of a nucleon at the point $\mathbf{r} = 0$ it is natural to give eq. (112.2) the form

$$\nabla^2 \psi - \kappa^2 \psi = -4\pi g \delta(\mathbf{r}). \quad (112.3)$$

We shall find the solution of this equation satisfying the condition $\psi \rightarrow 0$ for $r \rightarrow \infty$.

We seek ψ in the form

$$\psi(r) = \int \psi_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k}.$$

Then, using the expansion of the δ -function in a Fourier integral (see Vol. 1, Appendix III) and eq. (112.3), we find for $\psi_{\mathbf{k}}$ the value

$$\psi_{\mathbf{k}} = \frac{g}{2\pi^2} \frac{1}{k^2 + \kappa^2}.$$

For the field ψ we obtain an expression which is conveniently written in the form

$$\psi = \frac{g}{2\pi^2} \int \frac{e^{ikr \cos \theta} k^2 dk}{k^2 + \kappa^2} \sin \theta d\theta d\varphi.$$

The integration over the angles φ and θ gives

$$\psi = \frac{2g}{\pi r} \int_0^\infty \frac{k \sin kr}{k^2 + \kappa^2} dk. \quad (112.4)$$

In integrating this expression it is convenient to introduce the range of integration over k from $-\infty$ to $+\infty$. In this case we have

$$\psi = \frac{g}{\pi i r} \int_{-\infty}^\infty \frac{k e^{ikr}}{k^2 + \kappa^2} dk.$$

This integral is easily calculated by means of the theory of residues

$$\int_{-\infty}^{\infty} \frac{k e^{ikr}}{k^2 + \kappa^2} dk = 2\pi i \operatorname{Res}(k=i\kappa) = \pi i e^{-\kappa r},$$

whence we obtain the expression for the nuclear field potential

$$\psi = \frac{g}{r} e^{-\kappa r}, \quad (112.5)$$

called the Yukawa potential.

Formula (112.5) shows that the potential of nuclear forces decreases exponentially with increasing distance. The effective region in which ψ is different from zero has the size

$$R \approx \kappa^{-1} = \hbar/mc.$$

The size of this region is in order of magnitude the same as the range of nuclear forces determined experimentally.

For $m = 0$ the potential ψ goes over into the potential of the electrostatic field

$$\psi = g/r.$$

Thus the quantity g indeed plays in the Yukawa potential the same role as the charge e in the electrostatic potential, and can rightfully be called the nucleon charge. It should be stressed that the calculation carried out above can by no means pretend to be a quantitative characteristic of the nuclear force field.

In reality the interaction between nucleons is not of a static character. For a correct treatment of the processes of virtual π -meson exchange it is necessary to quantize the π -meson field ψ defined by the Klein—Gordon—Fock equation. This means that the function ψ and the adjoint function ψ^\dagger must be considered as quantum-mechanical operators in the space of occupation numbers. These operators have matrix elements different from zero for the processes of absorption and emission of π -mesons. The interaction between nucleons should be calculated by methods analogous to those applied in radiation theory.

We have seen in ch. 12 that the mathematical apparatus of radiation theory is based on the application of perturbation theory. The dimensionless interaction constant $e^2/\hbar c$, made up of the charge of the particle and of the universal constants \hbar and c , figures as the small parameter. The strong nuclear interaction can also be characterized by the interaction constant $g^2/\hbar c$. However, and here lies the profound difference from the electromagnetic

interaction, the quantity $g^2/\hbar c$ is of the order of ten. Thus the effectiveness of the nuclear interaction exceeds that of the electromagnetic interaction by a factor of more than a thousand. The term 'strong nuclear interaction' is associated with this fact. The large value of the interaction constant $g^2/\hbar c$ makes it impossible to use the apparatus of perturbation theory for the calculation of nuclear interactions.

This fact reflects the change in the physical nature of the interaction in passing from charged particles to nucleons. The smallness of the electromagnetic interaction constant means that the probability of emission of N particles in one act is proportional to $(e^2/\hbar c)^N \ll 1$. In other words, the probability of emission of one (actual or virtual) photon is considerably higher than that of the simultaneous emission of two, three and so on photons.

The situation is different in the case of the strong nuclear interaction. The probability of simultaneous emission of a large number of mesons is of the same order of magnitude as the probability of emission of one meson.

Hence each nucleon should be considered as a particle surrounded by a cloud of virtual π -mesons.

The validity of such a picture is confirmed by the phenomena of multiple production of π -mesons in collisions of high-energy nucleons.

Thus the picture of the π -meson interaction of nucleons turns out to be much more complex than that of the photon interaction of charges. The interaction between two nucleons involves without fail a multitude of π -mesons and its consideration must be based on the solution of the many-body problem. No consequential quantitative theory of the strong nuclear interaction has as yet been developed.

§113. The Dirac equation

In the preceding sections we have considered the relativistically invariant wave equation valid for spin zero particles. We have seen that the quantity ρ/e , which should be interpreted as the probability density, can take on negative as well as positive values.

As can be seen from formula (111.3), this is associated with the fact that the value of ρ/e is determined not only by the initial value of the ψ -function but also by the initial value of the derivative $\partial\psi/\partial t$, defined arbitrarily. It is clear that in order to eliminate this difficulty it is necessary to eliminate the possibility of an arbitrary choice of the derivative $\partial\psi/\partial t$. In other words, it is necessary that the relativistic generalization of the Schrödinger equation

contain only the first derivative with respect to time, as does the Schrödinger equation itself. Since, however, all relativistically invariant expressions must involve coordinates and time in the same way, the relativistic generalization of the Schrödinger equation should also involve first derivatives with respect to the coordinates.

The principle of superposition requires that the relativistic wave equation be linear. On the basis of these considerations Dirac formulated the following equation for the description of the motion of a free particle:

$$i\hbar \frac{\partial \psi}{\partial t} = \left(\beta'_x \frac{\partial}{\partial x} + \beta'_y \frac{\partial}{\partial y} + \beta'_z \frac{\partial}{\partial z} + \beta_0 \right) \psi. \quad (113.1)$$

Expression (113.1) represents the most general linear form containing only the first derivatives of the function sought. This equation is conveniently rewritten in a somewhat different form by redefining the quantities β' . Namely, we write it in the form

$$i\hbar \frac{\partial \psi}{\partial t} = (\hat{\beta}_x \hat{p}_x + \hat{\beta}_y \hat{p}_y + \hat{\beta}_z \hat{p}_z + \hat{\beta}_0) \psi,$$

where the operators \hat{p}_x , \hat{p}_y , \hat{p}_z are the ordinary operators of the components of the momentum along the coordinate axes, and the operators $\hat{\beta}_x$, $\hat{\beta}_y$, $\hat{\beta}_z$, $\hat{\beta}_0$ contain no coordinates. We determine the properties of these operators from the following reasoning. Introducing the notation

$$\hat{H} = \hat{\beta}_x \hat{p}_x + \hat{\beta}_y \hat{p}_y + \hat{\beta}_z \hat{p}_z + \hat{\beta}_0,$$

eq. (113.1) can be written in the form

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H} \psi, \quad (113.2)$$

which is completely, although as yet only formally, similar to the Schrödinger equation.

If it is assumed that the operator \hat{H} indeed represents a Hamiltonian, then there must be the same relation between \hat{H} and the momentum operators as between the energy and momentum in the theory of relativity, i.e.

$$\hat{H}^2 = c^2(\hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2) + m^2 c^4. \quad (113.3)$$

This requirement allows one to define the operators $\hat{\beta}_x$, $\hat{\beta}_y$, $\hat{\beta}_z$, $\hat{\beta}_0$. Indeed, squaring the operator \hat{H} we obtain

$$\begin{aligned}
\hat{H}^2 = & \hat{\beta}_x^2 \hat{p}_x^2 + \hat{\beta}_y^2 \hat{p}_y^2 + \hat{\beta}_z^2 \hat{p}_z^2 + \hat{\beta}_0^2 + \\
& + (\hat{\beta}_x \hat{\beta}_y + \hat{\beta}_y \hat{\beta}_x) \hat{p}_x \hat{p}_y + (\hat{\beta}_y \hat{\beta}_z + \hat{\beta}_z \hat{\beta}_y) \hat{p}_y \hat{p}_z + (\hat{\beta}_x \hat{\beta}_z + \hat{\beta}_z \hat{\beta}_x) \hat{p}_x \hat{p}_z + \\
& + (\hat{\beta}_x \hat{\beta}_0 + \hat{\beta}_0 \hat{\beta}_x) \hat{p}_x + (\hat{\beta}_y \hat{\beta}_0 + \hat{\beta}_0 \hat{\beta}_y) \hat{p}_y + (\hat{\beta}_z \hat{\beta}_0 + \hat{\beta}_0 \hat{\beta}_z) \hat{p}_z .
\end{aligned} \quad (113.4)$$

The operator \hat{H}^2 will have the form of (113.3) if the following relations are fulfilled:

$$\begin{aligned}
\hat{\beta}_x^2 = \hat{\beta}_y^2 = \hat{\beta}_z^2 = c^2, \quad \hat{\beta}_0^2 = m^2 c^4, \\
\hat{\beta}_i \hat{\beta}_k + \hat{\beta}_k \hat{\beta}_i = 0 \quad (i \neq k), \quad \hat{\beta}_i \hat{\beta}_0 + \hat{\beta}_0 \hat{\beta}_i = 0.
\end{aligned}$$

Here i and k take on the values x, y, z .

Usually in place of the operators $\hat{\beta}_i$ one introduces operators α_i which differ from the former by constant factors:

$$\hat{\beta}_x = c\alpha_x; \quad \hat{\beta}_y = c\alpha_y; \quad \hat{\beta}_z = c\alpha_z; \quad \hat{\beta}_0 = mc^2\beta$$

The following relations obviously hold for the operators α and β :

$$\begin{aligned}
\alpha_x^2 = \alpha_y^2 = \alpha_z^2 = \beta^2 = 1, \quad \alpha_i \beta + \beta \alpha_i = 0, \\
\alpha_i \alpha_k + \alpha_k \alpha_i = 0 \quad (i \neq k).
\end{aligned} \quad (113.5)$$

By means of these operators eq. (113.1) can be written in the form

$$i\hbar \frac{\partial \psi}{\partial t} = [c(\alpha_x \hat{p}_x + \alpha_y \hat{p}_y + \alpha_z \hat{p}_z) + mc^2 \beta] \psi. \quad (113.6)$$

This equation is called the Dirac equation.

If a vector operator is introduced by the equality

$$\boldsymbol{\alpha} = \alpha_x \mathbf{i} + \alpha_y \mathbf{j} + \alpha_z \mathbf{k},$$

then the Dirac equation can be written in a still more compact form

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H} \psi, \quad \hat{H} = c \boldsymbol{\alpha} \cdot \hat{\mathbf{p}} + mc^2 \beta. \quad (113.7)$$

We now seek the explicit form of the operators $\alpha_x, \alpha_y, \alpha_z, \beta$. We note, first of all, that the actions of these operators cannot reduce to the multiplication of the wave function by constant numbers. By means of such operators it would be impossible to satisfy relations (113.5).

Let us try to seek the operators $\alpha_x, \alpha_y, \alpha_z, \beta$ in the form of a set of constant, in general, complex numbers, i.e. in the form of the square matrices

$$\alpha_x = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

We first define the number n , which we assume to be the same for the matrices α and β . For this we make the matrices α and β correspond to the determinants

$$\det \alpha_x = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Before proceeding to further investigation of the matrices we note that the following relation must be fulfilled for the determinant of the product of the matrices:

$$\det \alpha_x \beta = \det \alpha_x \det \beta. \quad (113.8)$$

From the commutation rules it follows further that

$$\alpha_x \beta = -\beta \alpha_x = -I \beta \alpha_x.$$

Here I is the unit matrix. Making then use of relation (113.8) we find

$$\det \alpha_x \beta = \det \alpha_x \det \beta = \det (-I) \det \beta \det \alpha_x.$$

Since the determinants are ordinary numbers, we find that

$$\det (-I) = 1$$

and, consequently,

$$(-1)^n = 1. \quad (113.9)$$

Thus the number n must be even. If the number n were equal to two, then the matrices sought would be two-by-two matrices. We have already encountered such matrices in §59, where it was shown that there are 4 linearly independent two-by-two numerical matrices: 3 Pauli matrices and a unit

matrix. This last commutes with all the Pauli matrices and, consequently, does not satisfy the condition of anti-commutation (113.5). On the other hand, in the case of four-by-four matrices it turns out to be possible to construct matrices with the properties required. Namely, by a simple check it can be seen that the matrices

$$\begin{aligned}\alpha_x &= \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, & \alpha_y &= \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} \\ \alpha_z &= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, & \beta &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\end{aligned}\quad (113.10)$$

satisfy all the requirements formulated above. Matrices (113.10) can be written in an abbreviated form by making use of the Pauli matrices. Indeed, from definitions (60.14) and (60.15) and (113.10) it is clear that we have the relations

$$\begin{aligned}\alpha_x &= \begin{pmatrix} 0 & \sigma_x \\ \sigma_x & 0 \end{pmatrix}, & \alpha_y &= \begin{pmatrix} 0 & \sigma_y \\ \sigma_y & 0 \end{pmatrix}, \\ \alpha_z &= \begin{pmatrix} 0 & \sigma_z \\ \sigma_z & 0 \end{pmatrix}, & \beta &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.\end{aligned}\quad (113.11)$$

The matrices α and β are Hermitian matrices. This can be established by a simple check. If we transpose the matrices and carry out complex conjugation, then the matrices obtained will be the same as the original ones. Hence for these matrices one can write that $\alpha_x^\dagger = \alpha_x$, $\alpha_y^\dagger = \alpha_y$, $\alpha_z^\dagger = \alpha_z$ and $\beta^\dagger = \beta$.

If in place of the four-by-four matrices we had introduced matrices of a higher rank, then the formal set-up of the theory would not be violated. However, as will be clear from what follows, when the four-by-four matrices are introduced the general Dirac equation describes the properties of spin one-half particles.

Taking for α_x , α_y , α_z and β the matrix expression (113.10) in the form of four-by-four matrices, we have to assign four components to the wave function ψ . Indeed, only in this case do the 4 equations, into which the general

expression (113.7) resolves when four-by-four matrices are substituted into it, contain four unknown functions. The four-component function ψ (called the Dirac bispinor) can be written in the form of the matrix

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}.$$

We write down these equations in explicit form, making use of the rule of multiplication of matrices (see (45.6)):

$$i\hbar \frac{\partial \psi_1}{\partial t} = c(\hat{p}_x - i\hat{p}_y)\psi_4 + c\hat{p}_z\psi_3 + mc^2\psi_1,$$

$$i\hbar \frac{\partial \psi_2}{\partial t} = c(\hat{p}_x + i\hat{p}_y)\psi_3 - c\hat{p}_z\psi_4 + mc^2\psi_2,$$

$$i\hbar \frac{\partial \psi_3}{\partial t} = c(\hat{p}_x - i\hat{p}_y)\psi_2 + c\hat{p}_z\psi_1 - mc^2\psi_3,$$

$$i\hbar \frac{\partial \psi_4}{\partial t} = c(\hat{p}_x + i\hat{p}_y)\psi_1 - c\hat{p}_z\psi_2 - mc^2\psi_4.$$

The Dirac equation is easily generalized to the case of the motion of a charged particle in an external electromagnetic field. Namely, replacing the momentum operator $\hat{\mathbf{p}}$ by the operator $\hat{\mathbf{p}} - (e/c)\mathbf{A}$ according to the usual scheme and adding the operator $e\varphi$ to the operator \hat{H} , where \mathbf{A} and φ are the vector potential and scalar potential of the electromagnetic field, we obtain the Dirac equation

$$i\hbar \frac{\partial \psi}{\partial t} = \left[c\boldsymbol{\alpha} \left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A} \right) + e\varphi + mc^2\beta \right] \psi. \quad (113.12)$$

Let us bring the Dirac equation into a more symmetrical form. Multiplying eq. (113.7) by the operator β from the left-hand side we have

$$i\hbar\beta \frac{\partial \psi}{\partial t} = (c\beta\boldsymbol{\alpha}\mathbf{p} + mc^2\beta^2)\psi. \quad (113.13)$$

We now introduce the following system of matrices:

$$\begin{aligned} \gamma_1 &= -i\beta\alpha_x, & \gamma_2 &= -i\beta\alpha_y, \\ \gamma_3 &= -i\beta\alpha_z, & \gamma_4 &= \beta. \end{aligned} \quad (113.14)$$

It is easy to verify that the commutation rules for the matrices γ_i are the same as for the matrices α and β , viz.

$$\gamma_i \gamma_k + \gamma_k \gamma_i = 2\delta_{ik}.$$

By means of the matrices γ_i we can rewrite the Dirac equation in the form:

$$\sum_{i=1}^3 \gamma_i \frac{\partial \psi}{\partial x_i} + \frac{mc}{\hbar} \psi + \frac{\gamma_4}{ic} \frac{\partial \psi}{\partial t} = 0. \quad (113.15)$$

If we now introduce the coordinate $x_4 = ict$ we can transform the Dirac equation into a highly symmetrical form:

$$\gamma_\mu \frac{\partial \psi}{\partial x_\mu} + \frac{mc}{\hbar} \psi = 0. \quad (113.16)$$

With the aid of the operators $\hat{\mathbf{p}}_\mu$ and A_μ the last equation can be rewritten in the form

$$\left[\gamma_\mu \left(\hat{\mathbf{p}}_\mu - \frac{e}{c} A_\mu \right) - imc \right] \psi = 0. \quad (113.17)$$

§114. The probability density and probability current in Dirac's theory

We shall show, first of all, that the difficulty of interpreting the probability density ρ/e which we have encountered in discussing the Klein—Gordon—Fock equation is absent in the Dirac equation. Following the usual scheme, we write in addition to the Dirac equation

$$i\hbar \frac{\partial \psi}{\partial t} = (-i\hbar c \boldsymbol{\alpha} \nabla + mc^2 \beta) \psi, \quad (114.1)$$

the adjoint equation

$$-i\hbar \frac{\partial \psi^\dagger}{\partial t} = i\hbar c \nabla \psi^\dagger \boldsymbol{\alpha}^\dagger + mc^2 \psi^\dagger \beta^\dagger. \quad (114.2)$$

Here we have made use of the rule of conjugation of the product of matrices

$$(ab)^\dagger = b^\dagger a^\dagger.$$

Since the operators $\boldsymbol{\alpha}$ and β are Hermitian, then $\boldsymbol{\alpha}^\dagger = \boldsymbol{\alpha}$, $\beta^\dagger = \beta$, and we obtain

$$-i\hbar \frac{\partial \psi^\dagger}{\partial t} = i\hbar c \nabla \psi^\dagger \boldsymbol{\alpha} + mc^2 \psi^\dagger \beta. \quad (114.3)$$

We multiply eq. (114.1) by ψ^\dagger on the left, and eq. (114.3) by ψ on the right and subtract the second equation from the first. We have

$$i\hbar \left(\psi^\dagger \frac{\partial \psi}{\partial t} + \frac{\partial \psi^\dagger}{\partial t} \psi \right) = -i\hbar c [\psi^\dagger \boldsymbol{\alpha} \nabla \psi + (\nabla \psi^\dagger \boldsymbol{\alpha}) \psi] . \quad (114.4)$$

The parenthesis on the right-hand side of eq. (114.4) means that the gradient acts only on the function ψ^\dagger . The expression standing in the bracket can easily be transformed by means of the formula

$$\psi^\dagger (\boldsymbol{\alpha} \nabla) \psi + (\nabla \psi^\dagger \boldsymbol{\alpha}) \psi = \psi^\dagger \nabla \boldsymbol{\alpha} \psi + (\nabla \psi^\dagger \boldsymbol{\alpha}) \psi = \nabla (\psi^\dagger \boldsymbol{\alpha} \psi) .$$

Equation (114.4) is then written in the form

$$\frac{\partial \psi^\dagger \psi}{\partial t} = -c \nabla (\psi^\dagger \boldsymbol{\alpha} \psi) . \quad (114.5)$$

Comparing the expression obtained with the general formula (7.5) we see that the essentially positive quantity $\psi^\dagger \psi = \psi_1^* \psi_1 + \psi_2^* \psi_2 + \psi_3^* \psi_3 + \psi_4^* \psi_4$ represents the probability density. The vector defined by the equality $\mathbf{j} = c \psi^\dagger \boldsymbol{\alpha} \psi$ gives the probability current for a particle with the wave function ψ .

Thus, as in the Schrödinger theory, the wave function allows the usual probabilistic interpretation. From the linearity of the Dirac equation and the probabilistic interpretation of the function ψ it follows that the basic propositions of quantum mechanics remain valid: (1) the interpretation of the quantity $|c_m(t)|^2$, where $c_m(t)$ is the coefficient of the expansion

$$\psi = \sum_m c_m \psi_m$$

and ψ_m is the eigenfunction of a certain operator, as the probability of measuring the corresponding eigenvalue, (2) the definition of the mean value

$$\bar{L} = \int \psi^\dagger \hat{L} \psi dV .$$

Consequently, the entire structure of quantum mechanics also remains valid.

§ 115. The solution of the Dirac equation for a free particle

As the simplest example of the solution of the Dirac equation let us consider the motion of a free particle. We shall seek the solution of the Dirac equation for a freely moving particle in the usual way

$$i\hbar \frac{\partial \psi}{\partial t} = (c \boldsymbol{\alpha} \cdot \hat{\mathbf{p}} + mc^2 \beta) \psi . \quad (115.1)$$

Substituting the wave function $\psi = \psi_0 e^{-iEt/\hbar}$ into (115.1) we obtain the equation for the time-independent wave function ψ_0

$$E\psi_0 = (c\boldsymbol{\alpha} \cdot \hat{\mathbf{p}} + mc^2\beta)\psi_0. \quad (115.2)$$

We consider further states with a definite momentum and seek the solution of eq. (115.2) in the form of a plane wave

$$\psi_0 = u e^{i(\mathbf{p} \cdot \mathbf{r})/\hbar}.$$

Then for the function u we obtain the equation

$$Eu = (c\boldsymbol{\alpha} \cdot \mathbf{p} + mc^2\beta)u. \quad (115.3)$$

We write u in the form

$$u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} = \begin{pmatrix} w \\ w' \end{pmatrix}, \quad w = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad w' = \begin{pmatrix} u_3 \\ u_4 \end{pmatrix}. \quad (115.4)$$

Substituting (115.4) into (115.3) and taking into account the representations of matrices α and β (113.11), we find

$$Ew = c\boldsymbol{\sigma} \cdot \mathbf{p}w' + mc^2w, \quad (115.5)$$

$$Ew' = c\boldsymbol{\sigma} \cdot \mathbf{p}w - mc^2w'. \quad (115.6)$$

Each of the functions w and w' has two components.

For the system of linear equations obtained to have a solution it is necessary that its determinant reduce to zero

$$\begin{vmatrix} E - mc^2 & -c\boldsymbol{\sigma} \cdot \mathbf{p} \\ -c\boldsymbol{\sigma} \cdot \mathbf{p} & E + mc^2 \end{vmatrix} = 0. \quad (115.7)$$

Evaluating the determinant we obtain

$$E^2 - m^2c^4 = +c^2(\boldsymbol{\sigma} \cdot \mathbf{p})^2.$$

The expression $(\boldsymbol{\sigma} \cdot \mathbf{p})^2$ can easily be transformed by means of the known properties of the Pauli matrices. According to (60.17) we have

$$(\boldsymbol{\sigma} \cdot \mathbf{p})^2 = p^2. \quad (115.8)$$

As was to be expected, we arrive at the relation already known between the energy and momentum of a particle

$$E^2 = c^2p^2 + m^2c^4. \quad (115.9)$$

The energy of a particle can take on positive as well as negative values. We have already discussed this problem in the theory of relativity and have seen that within the framework of classical mechanics this fact did not lead to any difficulties, since the energy range of width $2mc^2$ is forbidden. Indeed, in classical mechanics all variables change continuously and a particle has either a positive or a negative energy. Continuous transition from one region into the other is impossible.

In relativistic quantum mechanics there are no grounds for rejecting the negative sign. We shall discuss in detail the meaning of the negative sign of the energy later.

Choosing the plus or minus sign for the energy we can solve the system of eqs. (115.5) and (115.6). By virtue of the homogeneity of the system of equations one of the quantities, either w or w' , remains arbitrary.

Let w be an arbitrary quantity. Then

$$w' = \frac{c \boldsymbol{\sigma} \cdot \mathbf{p}}{E + mc^2} w. \quad (115.10)$$

If, on the contrary, the quantity w' is assumed to be arbitrary, then we have

$$w = \frac{c \boldsymbol{\sigma} \cdot \mathbf{p}}{E - mc^2} w'. \quad (115.11)$$

The corresponding wave functions are of the form (for simplicity the direction of the momentum vector is here taken to be the z -axis)

$$u = \begin{bmatrix} A \\ B \\ \frac{cp_z A}{E + mc^2} \\ -\frac{cp_z B}{E + mc^2} \end{bmatrix}, \quad v = \begin{bmatrix} \frac{cp_z D}{E - mc^2} \\ \frac{cp_z F}{E - mc^2} \\ D \\ F \end{bmatrix} \quad (115.12)$$

Here A, B, D and F are arbitrary constants. The character of these expressions will become clearer if we pass to the non-relativistic limit, setting $E \sim mc^2$ or $E \sim -mc^2$ respectively. Then from (115.10) it is seen that in the first case

$$w' = \frac{c \boldsymbol{\sigma} \cdot \mathbf{p}}{mc^2} w \sim \frac{v}{c} w.$$

The spinor w' is less than w in the ratio v/c , so that $u \approx \begin{bmatrix} A \\ B \\ 0 \\ 0 \end{bmatrix}$.

For the negative value of energy (115.11) gives

$$w \approx -\frac{c \boldsymbol{\sigma} \cdot \mathbf{p} w'}{2mc^2} \approx -\frac{v}{c} w' \quad \text{and} \quad v = \begin{bmatrix} 0 \\ 0 \\ D \\ F \end{bmatrix}.$$

Thus in the transition to the non-relativistic approximation two components of the wave function turn out to be small in comparison with two other components. In this case for positive energies w is large in comparison with w' , and for negative energies the reverse is true.

The general solution of the Dirac equation for the motion of a free particle can be written in the form of a superposition of wave functions of the type (115.12), i.e. as a Fourier integral of the form

$$\psi(x, y, z, t) = \int u(A, B) e^{-iEt/\hbar} e^{i(\mathbf{p} \cdot \mathbf{r})/\hbar} d\mathbf{p} + \int v(D, F) e^{+iEt/\hbar} e^{i(\mathbf{p} \cdot \mathbf{r})/\hbar} d\mathbf{p},$$

where $d\mathbf{p} = dp_x dp_y dp_z$.

If at the initial instant of time $t = 0$ the following wave function is given

$$\psi(x, y, z, 0) = \int \varphi(\mathbf{p}) e^{i(\mathbf{p} \cdot \mathbf{r})/\hbar} d\mathbf{p}; \quad \varphi(\mathbf{p}) = \begin{bmatrix} \varphi_1(\mathbf{p}) \\ \varphi_2(\mathbf{p}) \\ \varphi_3(\mathbf{p}) \\ \varphi_4(\mathbf{p}) \end{bmatrix},$$

then a given set of quantities $\varphi_1, \varphi_2, \varphi_3$ and φ_4 can be defined unambiguously in terms of the four arbitrary coefficients involved in u and v .

Thus a set of two waves, one of which corresponds to the positive energy and the other to the negative energy, forms the total solution of the Dirac equation. It is clear that if the particular solution corresponding to the negative energy were rejected and only the solution with the positive energy were retained, then the system of functions found in this case would be incomplete.

The initial conditions contain four given quantities, whereas u involves only two indeterminate constants A and B . Thus irrespective of other considerations the necessity of taking into account solutions with negative energy follows from the general foundations of quantum mechanics.

In the next section we shall come back to the discussion of the fundamental conclusions which have been drawn from the existence of solutions of the Dirac equation which correspond to a negative energy of the particle.

§116. The concept of the positron

We now turn to the discussion of the formula

$$E = \pm(p^2c^2 + m^2c^4)^{\frac{1}{2}}. \quad (116.1)$$

As has already been pointed out, from the point of view of classical mechanics the negative energy of a free particle has no physical meaning.

In quantum mechanics the situation is different. Namely, discontinuous transitions are possible from states with a positive energy into states with a negative energy. In other words, these two classes of states are no longer separated by an impenetrable barrier. We have already seen that the exclusion of states with a negative energy contradicts the general propositions of quantum mechanics, since the wave functions of states with a positive energy do not form a complete system of functions.

On the other hand, it is impossible to assume the existence of particles with a negative energy. Such particles would possess properties which differ fundamentally from those of all particles observed in nature. As an example we can point to the following: a particle with a negative energy $-|E_1|$ could make a transition into a state with a lower negative energy $-|E_2|$, $|E_2| > |E_1|$. Then the difference $|E_2| - |E_1|$ could be converted into useful work. Such a transition could be carried out continuously, since $|E_2|$ is in no way limited, and a particle with a negative energy could serve as an infinitely large source of work.

In order to avoid difficulties associated with the introduction of observable particles with negative energy into the theory, Dirac introduced the concept of the vacuum as that state of space in which all states with negative energy are occupied by electrons and all states with positive energy are free. According to the Pauli principle, there is one electron in each state with a negative energy.

We assume further that under the influence of an external action one of the electrons is removed from a state with a negative energy. The vacant state with negative energy manifests itself as 'something' with a positive energy, since for the destruction of such a state, i.e. for its occupation it is necessary to add to it an electron with a negative energy. Thus the vacant state with a negative energy should be treated as a particle having a positive energy.

It should be noted that Dirac at first incorrectly assumed that this state corresponded to the proton. Subsequently it was shown theoretically that the particle corresponding to a vacant state with a negative energy must have a mass equal to the mass of the electron and, consequently, it could not be the proton.

Let us consider in more detail the considerations of Dirac about the occupied background of negative energies. Let $N_{\alpha}^{(-)}(p)$ and $N_{\alpha}^{(+)}(p)$ denote the numbers of electrons which are respectively in states with a negative and a positive energy and have momentum p and a definite spin orientation. The index α can take on two values according to the direction of spin. These numbers, in accordance with the Pauli principle, can take on only the values 0 or 1. In the vacuum state (the index v) we have

$$N_{\alpha v}^{(-)}(p) = 1, \quad N_{\alpha v}^{(+)}(p) = 0$$

for all momentum values.

Indeed, all states with negative energy are then occupied, and all states with positive energy are vacant. The energy E_v and charge q_v in the vacuum are defined by the relations

$$E_v = - \sum_{\alpha, p} E(p) N_{\alpha v}^{(-)}(p), \quad (116.2)$$

$$q_v = - |e| \sum_{\alpha, p} N_{\alpha v}^{(-)}(p). \quad (116.3)$$

Here e is the charge of the electron.

Since the momentum and energy of free particles are in no way limited, the values of E_v and q_v are infinitely large. However, according to Dirac, these quantities are in principle not observable. Only those quantities which characterize the departure from the vacuum state are observable.

Further, we write the total energy E of the system and the charge q of the system in the case where there are in space electrons in states with positive energy while there are vacancies in states with negative energy

$$E = \sum_{\alpha, p} [N_{\alpha}^{(+)}(p) - N_{\alpha}^{(-)}(p)] E(p), \quad (116.4)$$

$$q = - \sum_{\alpha, p} |e| [N_{\alpha}^{(+)}(p) + N_{\alpha}^{(-)}(p)]. \quad (116.5)$$

In correspondence with the above only the following differences are observable:

$$E - E_v = \sum_{\alpha, p} [N_{\alpha}^{(+)}(p) + (N_{\alpha v}^{(-)}(p) - N_{\alpha}^{(-)}(p))] E(p), \quad (116.6)$$

$$q - q_v = -|e| \sum_{\alpha, p} [N_{\alpha}^{(+)}(p) + N_{\alpha}^{(-)}(p) - N_{\alpha v}^{(-)}(p)]. \quad (116.7)$$

From formulae (116.6) and (116.7) we see that if a certain state with a negative energy is vacant, i.e. $N_{\alpha}^{(-)}(p) = 0$, then it corresponds to a positive contribution to the observed values of energy and charge. Indeed, formulae (116.6) and (116.7) involve the expressions $N_{\alpha v}^{(-)}(p) - N_{\alpha}^{(-)}(p)$. If a state with a negative energy is vacant $N_{\alpha}^{(-)}(p) = 0$, then $N_{\alpha v}^{(-)}(p) - N_{\alpha}^{(-)}(p) = 1$. In this case a positive contribution to the energy and charge of the system arises, equal respectively to E_p and $|e|$. Thus we see that the absence of an electron with momentum p in the continuous background of occupied negative states is equivalent to the appearance of an observable particle with a positive charge, positive energy and momentum $-p$. Such a particle, with charge $(+|e|)$ and a mass equal to the mass of the electron, was called a positron. It was discovered by Anderson in cosmic rays a few years after the appearance of Dirac's theory.

Proceeding from Dirac's concepts it turns out to be possible to account for a number of known physical effects. For example, it is obvious that an electromagnetic field can produce an electron-positron pair if the energy of the photon $\hbar\omega$ is greater than $2mc^2$. This energy is necessary for bringing an electron from a state with a negative energy into a state with a positive energy.

The laws of conservation of energy and momentum restrict the possibility of the reaction of electron-positron pair production by a photon. In fact this reaction can take place only in the presence of a third body — for example a nucleus, which takes a part of the momentum. In addition to electron-positron pair production the inverse reaction, positron annihilation, is possible. In the annihilation an electron with a positive energy makes a transition into a vacant state with negative energy. The difference between the energies is emitted in the form of a γ -quantum.

Dirac's theory made it possible to predict not only these phenomena but also to calculate the cross section for both processes. The excellent agreement of the results of the calculations with experimental data was a strong confirmation of the validity of Dirac's ideas. However, the last decade has been marked by very important theoretical and experimental achievements, which will in part be elucidated in what follows. These successes allow one, on the

one hand, to show the reality of the existence of the vacuum in the Dirac sense, and, on the other hand, to extend the regions of applicability of relativistic quantum mechanics. As has already been pointed out, the vacuum represents a system of charged particles occupying all possible states. When an external electric charge is brought into the vacuum, or when an electromagnetic field arises, the vacuum begins to interact with the external fields. For example, Lamb discovered in 1953 that the levels $^2S_{\frac{1}{2}}$ and $^2P_{\frac{1}{2}}$ of the hydrogen atom have somewhat different energies (the Lamb shift). This effect can be accounted for only by an interaction with the vacuum (see §119 and §128).

Thus the concept of the vacuum developed by Dirac is confirmed by a number of diverse experiments.

The symmetry of the theory with respect to electrons and positrons finds its expression in the fact that there exists a unitary operator \hat{C} , called the charge conjugation operator, which transforms a particle into its antiparticle. In other words, the action of the operator \hat{C} exchanges an electron and a positron (of the same spin and energy).

If ψ_e and ψ_p denote respectively the wave functions of an electron and of a positron, then by definition we can write for them (see (113.17))

$$\left[\gamma_\mu \left(\hat{p}_\mu - \frac{e}{c} A_\mu \right) - imc \right] \psi_e = 0, \quad (116.8)$$

$$\left[\gamma_\mu \left(\hat{p}_\mu + \frac{e}{c} A_\mu \right) - imc \right] \psi_p = 0. \quad (116.9)$$

Then the function ψ_e^* which is complex-conjugate to ψ_e satisfies the equation

$$\left[\gamma_\mu^* \left(\hat{p}_\mu^* - \frac{e}{c} A_\mu^* \right) + imc \right] \psi_e^* = 0,$$

or, since

$$\hat{p}_i^* = -\hat{p}_i, \quad A_i^* = A_i \quad (i=1,2,3), \quad \hat{p}_4^* = \hat{p}_4, \quad A_4^* = -A_4,$$

we have

$$\left[-\gamma_4^* \left(\hat{p}_4 + \frac{e}{c} A_4 \right) + \gamma_i^* \left(\hat{p}_i + \frac{e}{c} A_i \right) - imc \right] \psi_e^* = 0. \quad (116.10)$$

From the comparison of (116.9) and (116.10) it is natural to assume that

$$\psi_p = \hat{C} \psi_e^*, \quad \psi_e^* = \hat{C}^{-1} \psi_p. \quad (116.11)$$

Substituting (116.11) into (116.9) we have

$$\left[\gamma_i \left(\hat{p}_i + \frac{e}{c} A_i \right) + \gamma_4 \left(\hat{p}_4 + \frac{e}{c} A_4 \right) - imc \right] \hat{C} \psi_e^* = 0.$$

Multiplying this equation from the left by \hat{C}^{-1} , we have

$$\hat{C}^{-1} \left[\gamma_i \left(\hat{p}_i + \frac{e}{c} A_i \right) + \gamma_4 \left(\hat{p}_4 + \frac{e}{c} A_4 \right) - imc \right] \hat{C} \psi_e^* = 0,$$

or

$$\left[\left(\hat{p}_i + \frac{e}{c} A_i \right) \hat{C}^{-1} \gamma_i \hat{C} + \left(\hat{p}_4 + \frac{e}{c} A_4 \right) \hat{C}^{-1} \gamma_4 \hat{C} - imc \right] \psi_e^* = 0. \quad (116.12)$$

For (116.12) to be identical with (116.10) it is necessary that the following equalities be fulfilled:

$$\hat{C}^{-1} \gamma_i \hat{C} = \gamma_i^*, \quad \hat{C}^{-1} \gamma_4 \hat{C} = -\gamma_4^* = -\gamma_4.$$

If γ_i and γ_4 are defined by formulae (113.14), then

$$\hat{C} = \gamma_2 = \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{vmatrix}.$$

From definition (116.11) it is seen directly that the operator \hat{C} commutes with the Hamiltonian. Thus one can introduce two wave functions ψ_e and ψ_p which are completely equivalent and interrelated by the following relations conserved in time:

$$\psi_p = \hat{C} \psi_e^* = \gamma_2 \psi_e^*, \quad \psi_e = \hat{C}^{-1} \psi_p^* = \gamma_2^{-1} \psi_p^*.$$

The two wave functions describe particles of the same (positive) energy, mass and spin, but with different signs of the charge and of the magnetic moment. The introduction of charge-conjugate wave functions for equivalent particles removes to a certain extent the logical difficulties associated with the simplified interpretation of the vacuum as a background filled by particles of negative energy.

In following chapters we shall describe in more detail the modern theory of fields and of elementary particles.

§117. The spin of particles described by the Dirac equation

Although we have so far used the concept of the spin of a particle extensively, the spin operator was introduced in a purely formal way, as a necessary tool for the description of experimental data. We shall now show that the existence of spin follows directly from the Dirac equation. For this we shall consider the conservation laws resulting from the Dirac equation.

Since in Dirac's theory all the general propositions of quantum mechanics are preserved, to find the conservation laws it is only necessary to find the commutator with the Hamiltonian. The difference from Schrödinger's theory lies in the fact that the Hamiltonian now has the form of (113.7).

If the Hamiltonian does not depend on time (and for this it is necessary that the external field potentials be time-independent), then the energy conservation law holds. In this respect there is no difference between the Schrödinger and the Dirac theories.

For a particle moving in the vacuum the total angular momentum must also be conserved. Hence there must exist a total angular momentum operator commuting with the Hamiltonian.

An interesting result is obtained when the operator for the orbital angular momentum operator $\hat{\mathbf{L}} = \hat{\mathbf{r}} \times \hat{\mathbf{p}}$ is commuted with the Hamiltonian.

For our purposes we restrict ourselves to the case of a free particle. Choosing an arbitrarily oriented z -axis, we have

$$\hat{H}\hat{L}_z - \hat{L}_z\hat{H} = (c\boldsymbol{\alpha} \cdot \mathbf{p} + mc^2\beta)\hat{L}_z - \hat{L}_z(c\boldsymbol{\alpha} \cdot \mathbf{p} + mc^2\beta).$$

Since the operator

$$\hat{L}_z = \frac{\hbar}{i} \left(y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \right)$$

commutes with the operators $\hat{\beta}$ and $\hat{\alpha}_z\hat{p}_z$, we then obtain

$$\hat{H}\hat{L}_z - \hat{L}_z\hat{H} = c\alpha_x(\hat{p}_x\hat{L}_z - \hat{L}_z\hat{p}_x) + c\alpha_y(\hat{p}_y\hat{L}_z - \hat{L}_z\hat{p}_y). \quad (117.1)$$

Making use of the property of commutation of the momentum components with the angular momentum components, we find

$$\hat{H}\hat{L}_z - \hat{L}_z\hat{H} = i\hbar c(\alpha_y\hat{p}_x - \alpha_x\hat{p}_y). \quad (117.2)$$

We obtain analogous results for the other momentum components.

Thus the orbital angular momentum is not a constant of the motion and is not conserved. To find the quantity playing the role of the total angular momentum we introduce the operator $\hat{\mathbf{J}} = \hat{\mathbf{L}} + \hat{\mathbf{S}}$, where $\hat{\mathbf{S}}$ is an unknown operator. We require that the operator $\hat{\mathbf{J}}$ commute with the operator \hat{H} :

$$\hat{H}\hat{J}_i - \hat{J}_i\hat{H} = 0 \quad \text{or} \quad (\hat{H}\hat{L}_i - \hat{L}_i\hat{H}) + (\hat{H}\hat{s}_i - \hat{s}_i\hat{H}) = 0.$$

Substituting $i = z$ and using the value of the commutator (117.2), we have

$$\hat{H}\hat{s}_z - \hat{s}_z\hat{H} = i\hbar c(\alpha_x\hat{p}_y - \alpha_y\hat{p}_x). \quad (117.3)$$

We try to satisfy this equation, setting

$$\hat{s}_z = A\alpha_x\alpha_y, \quad (117.4)$$

where A is an unknown constant.

We further calculate the commutator $\hat{H}\hat{s}_z - \hat{s}_z\hat{H}$. Making use of (113.5), we obtain

$$\begin{aligned} A\hat{H}\alpha_x\alpha_y - A\alpha_x\alpha_y\hat{H} &= \\ &= Ac(\alpha_x\hat{p}_x + \alpha_y\hat{p}_y)\alpha_x\alpha_y - Ac\alpha_x\alpha_y(\alpha_x\hat{p}_x + \alpha_y\hat{p}_y) = \\ &= 2Ac(\alpha_y\hat{p}_x - \alpha_x\hat{p}_y). \end{aligned}$$

Comparing this expression with formula (117.3), we find the value $A = -\frac{1}{2}i\hbar$. Thus the operator \hat{s}_z is equal to

$$\begin{aligned} \hat{s}_z &= -\frac{i\hbar}{2}\alpha_x\alpha_y = -\frac{i\hbar}{2}\begin{pmatrix} 0 & \sigma_x \\ \sigma_x & 0 \end{pmatrix}\begin{pmatrix} 0 & \sigma_y \\ \sigma_y & 0 \end{pmatrix} = -\frac{i\hbar}{2}\begin{pmatrix} \sigma_x\sigma_y & 0 \\ 0 & \sigma_x\sigma_y \end{pmatrix} = \\ &= -\frac{i\hbar}{2}\begin{pmatrix} i\sigma_z & 0 \\ 0 & i\sigma_z \end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix} \sigma_z & 0 \\ 0 & \sigma_z \end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \end{aligned} \quad (117.5)$$

The two other vector components \hat{s}_x and \hat{s}_y are obtained from analogous calculations

$$\hat{s}_x = -\frac{1}{2}i\hbar\alpha_y\alpha_z; \quad \hat{s}_y = -\frac{1}{2}i\hbar\alpha_z\alpha_x.$$

We now find the operator $\hat{s}^2 = \hat{s}_x^2 + \hat{s}_y^2 + \hat{s}_z^2$. Making use of the properties of the operators $\alpha_x, \alpha_y, \alpha_z$, we find

$$\hat{s}^2 = \frac{3}{4}\hbar^2\hat{I} = \hbar^2\frac{1}{2}(1 + \frac{1}{2})\hat{I}. \quad (117.6)$$

We now turn to the discussion of the results obtained. It is evident that the quantity \mathbf{J} , conserved in time, should be considered as the total angular momentum of the particle. In its turn the total angular momentum is the sum of the orbital and the intrinsic spin angular momenta of the particle. The operators \hat{s}_z and \hat{s}^2 in formulae (117.5) and (117.6) are brought to

diagonal form. The spin component along the z -axis can then take on the two values $\pm \frac{1}{2}\hbar$. The eigenvalues of the operator \hat{s}^2 have the form of $\hbar^2 s(s+1)$, where $s = \frac{1}{2}$. Hence it is obvious that the particle has spin $\frac{1}{2}\hbar$.

§118. The transition from the Dirac equation to the Pauli equation. The magnetic moment of a particle

Let us now see how the Dirac equation transforms if the transition to the non-relativistic approximation is carried out in it. We shall consider the general case where the particle moves in an external electromagnetic field, so that the Dirac equation has the form of (113.12). Just as in the limiting transition in the scalar relativistic equation, we first of all separate the rest energy, i.e. we carry out a transformation of the form

$$\psi = \psi' e^{-imc^2 t/\hbar}.$$

Then for the function ψ' we obtain the equation

$$i\hbar \frac{\partial \psi'}{\partial t} = \left[c\boldsymbol{\alpha} \cdot \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right) + mc^2(\beta - 1) + e\varphi \right] \psi'. \quad (118.1)$$

If the wave function is written in the form $\psi' = \begin{pmatrix} w \\ w' \end{pmatrix}$, then just as for a free particle we obtain the equations for w and w' :

$$\begin{aligned} i\hbar \frac{\partial w}{\partial t} &= c\boldsymbol{\sigma} \cdot \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right) w' + e\varphi w, \\ i\hbar \frac{\partial w'}{\partial t} &= c\boldsymbol{\sigma} \cdot \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right) w - 2mc^2 w' + e\varphi w'. \end{aligned} \quad (118.2)$$

As always, the limiting transition to the non-relativistic approximation corresponds to a formal expansion in powers of c . We assume at first that in the general case of the motion of a particle in a field, as well as for a free particle, $w' \sim c^{-1}w$. Then in the second of eqs. (118.2) we can disregard the terms $i\hbar(\partial w'/\partial t)$ and $e\varphi w'$, since they are small in comparison with the quantities

$$c\boldsymbol{\sigma} \cdot \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right) w \quad \text{and} \quad mc^2 w',$$

which are proportional to c . We then obtain for the spinor w' the expression

$$w' = \frac{1}{2mc} \boldsymbol{\sigma} \cdot \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right) w, \quad (118.3)$$

which is in agreement with our assumption.

Substituting (118.3) into the first of eqs. (118.2) we find

$$i\hbar \frac{\partial w}{\partial t} = \frac{\left[\boldsymbol{\sigma} \cdot \left(\mathbf{p} - \frac{e}{c} \mathbf{A} \right) \right]^2}{2m} w + e\varphi w. \quad (118.4)$$

We evaluate the square of the operator in the explicit form

$$\left[\boldsymbol{\sigma} \cdot \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right) \right]^2 = \left[\sigma_x \left(\hat{p}_x - \frac{e}{c} A_x \right) + \sigma_y \left(\hat{p}_y - \frac{e}{c} A_y \right) + \sigma_z \left(\hat{p}_z - \frac{e}{c} A_z \right) \right]^2.$$

In multiplying it should be recalled that the operators $\hat{\mathbf{p}}$ and \mathbf{A} do not commute with each other. Carrying out the multiplication we find

$$\begin{aligned} \left[\boldsymbol{\sigma} \cdot \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right) \right]^2 = & \sigma_x^2 \left(\hat{p}_x - \frac{e}{c} A_x \right)^2 + \sigma_y^2 \left(\hat{p}_y - \frac{e}{c} A_y \right)^2 + \sigma_z^2 \left(\hat{p}_z - \frac{e}{c} A_z \right)^2 \\ & + \sigma_x \sigma_y \left(\hat{p}_x - \frac{e}{c} A_x \right) \left(\hat{p}_y - \frac{e}{c} A_y \right) + \sigma_y \sigma_x \left(\hat{p}_y - \frac{e}{c} A_y \right) \left(\hat{p}_x - \frac{e}{c} A_x \right) \\ & + \sigma_x \sigma_z \left(\hat{p}_x - \frac{e}{c} A_x \right) \left(\hat{p}_z - \frac{e}{c} A_z \right) + \sigma_z \sigma_x \left(\hat{p}_z - \frac{e}{c} A_z \right) \left(\hat{p}_x - \frac{e}{c} A_x \right) \\ & + \sigma_y \sigma_z \left(\hat{p}_y - \frac{e}{c} A_y \right) \left(\hat{p}_z - \frac{e}{c} A_z \right) + \sigma_z \sigma_y \left(\hat{p}_z - \frac{e}{c} A_z \right) \left(\hat{p}_y - \frac{e}{c} A_y \right). \quad (118.5) \end{aligned}$$

According to (60.16) we have for the Pauli matrices $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$. We see that the sum of the first three terms is brought to the form

$$\sigma_x^2 \left(\hat{p}_x - \frac{e}{c} A_x \right)^2 + \sigma_y^2 \left(\hat{p}_y - \frac{e}{c} A_y \right)^2 + \sigma_z^2 \left(\hat{p}_z - \frac{e}{c} A_z \right)^2 = \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right)^2.$$

We carry out the further transformation only with the terms

$$\sigma_x \sigma_y \left(\hat{p}_x - \frac{e}{c} A_x \right) \left(\hat{p}_y - \frac{e}{c} A_y \right) + \sigma_y \sigma_x \left(\hat{p}_y - \frac{e}{c} A_y \right) \left(\hat{p}_x - \frac{e}{c} A_x \right), \quad (118.6)$$

since the remaining expressions transform in a way analogous to (118.6). The matrices σ_x and σ_y anticommute and, consequently, expression (118.6) can be rewritten in the form

$$\frac{e}{c} \sigma_x \sigma_y [-\hat{p}_x A_y - A_x \hat{p}_y + \hat{p}_y A_x + A_y \hat{p}_x]. \quad (118.7)$$

Making use of the commutation properties of the operators \hat{p}_x and \hat{p}_y with the operators depending on coordinates (26.10), we have

$$\begin{aligned} \frac{e}{c} \sigma_x \sigma_y \left[-i\hbar \frac{\partial A_x}{\partial y} + i\hbar \frac{\partial A_y}{\partial x} \right] &= \frac{ie\hbar}{c} \sigma_x \sigma_y \left[\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right] = \\ &= \frac{ie\hbar}{c} \sigma_x \sigma_y (\nabla_z \times \mathbf{A}) = \frac{ie\hbar}{c} \sigma_x \sigma_y \mathcal{H}_z. \end{aligned}$$

Since, according to (60.16), $\sigma_x \sigma_y = i\sigma_z$, then we finally have

$$\frac{ie\hbar}{c} \sigma_x \sigma_y \mathcal{H}_z = -\frac{e\hbar}{c} \sigma_z \mathcal{H}_z.$$

Carrying out analogous transformations with the remaining terms of (118.5) we obtain

$$\left[\boldsymbol{\sigma} \cdot \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right) \right]^2 = \left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right)^2 - \frac{e\hbar}{c} \boldsymbol{\sigma} \cdot \mathcal{H}. \quad (118.8)$$

Substituting (118.8) into (118.4) we find

$$i\hbar \frac{\partial w}{\partial t} = \left[\frac{\left(\hat{\mathbf{p}} - \frac{e}{c} \mathbf{A} \right)^2}{2m} + e\varphi - \frac{e\hbar}{2mc} \boldsymbol{\sigma} \cdot \mathcal{H} \right] w. \quad (118.9)$$

We see that in the transition to the non-relativistic approximation the Dirac equation automatically goes over into the Pauli equation. Hence from Dirac's theory it is seen that there results, not only the existence of the spin of particles (equal to $\frac{1}{2}\hbar$), but also the existence of the intrinsic magnetic moment of particles

$$\mu = \frac{e\hbar}{2mc}. \quad (118.10)$$

We can now define more precisely the problem as to what are the particles having spin $\frac{1}{2}\hbar$ to which the Dirac equation can be applied. If m is understood to be the mass of the electron, then good agreement is obtained between the calculated and measured values of the magnetic moment.

Thus the Dirac equation describes the behaviour of electrons with a high degree of accuracy. The Dirac equation also makes it possible, apparently, to describe well the properties of the neutrino, a particle with rest mass $m = 0$ (see §123).

However, attempts to apply the Dirac equation to heavy particles of spin $\frac{1}{2}$, the proton and the neutron, have not led to very satisfactory results. On

the other hand, it has also been possible to obtain some general and very important conclusions from the Dirac equation for heavy particles.

It turns out that the behaviour of fast protons and neutrons described qualitatively also fits the framework of the Dirac equation. Of particular importance is the fact that the basic idea of Dirac's theory, the existence of antiparticles, has received direct confirmation for mesons as well as for nucleons.

The antiproton \bar{p} , a particle with negative elementary charge and a mass equal to that of the proton, was discovered in 1955 in the reaction $p + p \rightarrow p + (\bar{p} + p) + p$ using an accelerator. Somewhat later the reaction $p + \bar{p} \rightarrow n + \bar{n}$ with the production of antineutrons was observed. The antineutron differs from the neutron in the sign of its magnetic moment and in its parity. When antiparticles are annihilated other particles are produced. For example, when protons and antiprotons are annihilated π^- and K-mesons are produced.

In spite of all these facts, quantitative calculations and, in particular, calculations of the magnetic moment are in disagreement with experimental data. If m in formula (118.10) is assumed to be the mass of the proton, then a value differing from the experimental value by a factor of 2.7 is obtained for its magnetic moment.

This disagreement of theory with experiment is apparently associated with the fact that heavy particles, protons and neutrons, interact strongly with the meson field. Herein lies their difference from electrons, which interact relatively weakly with the electromagnetic field*.

§ 119. The hydrogen atom in Dirac's theory

Although the motion of the electron in the hydrogen atom corresponds to non-relativistic velocities, finding the relativistic corrections to the hydrogen energy levels was of great interest, since Schrödinger's theory could not account for the appearance of fine structure in the hydrogen spectrum.

In §38 it was found that the energy levels of the hydrogen atoms depend only on the principal quantum number. However, experiment shows that the principal quantum number characterizes the energy levels only approximately. In reality the excited levels are split into close sub-levels. As a result a splitting of the spectral lines, clearly observable in an ordinary spectrometer and

* For more detailed considerations on the possibility of applying the Dirac equation to nucleons see A.I.Akhiezer and V.B.Berestetskii, *Quantum electrodynamics* (Interscience Publishers, New York, 1965).

particularly accurately measured by means of modern radiospectroscopic methods, was observed in the hydrogen spectrum. It turns out that this splitting of levels is associated with the spin-orbit interaction and that it follows from Dirac's theory.

The Dirac equation for stationary state motion in the Coulomb field is of the form

$$[c\boldsymbol{\alpha} \cdot \hat{\mathbf{p}} + mc^2\beta]\psi = \left(E + \frac{Ze^2}{r}\right)\psi.$$

The Dirac equation, as well as the Schrödinger equation, allows exact solution for the Coulomb field. However, in contrast to the Schrödinger equation, the Dirac equation does not lead to distinct laws of conservation of total angular momentum (see §117). Calculations show that only in the non-relativistic approximation can one speak of constant values of the orbital and spin angular momenta. In this case it turns out that the Hamiltonian assumes the form*

$$\hat{H} = \frac{\hat{\mathbf{p}}^2}{2m} + U + \frac{1}{2mc^2} \frac{1}{r} \frac{\partial U}{\partial r} (\hat{\mathbf{s}} \cdot \hat{\mathbf{L}}) + O\left(\frac{1}{c^2}\right), \quad (119.1)$$

where the first two terms are the same as the Hamiltonian of the Schrödinger equation, and the third term represents the spin-orbit interaction energy. Terms of the order of $1/c^2$, which are not written out, contain relativistic corrections to the kinetic and potential energies which do not have an obvious interpretation.

Solving the Dirac equation leads to the following expression for the energy of the electron:

$$E = mc^2 - \frac{Z^2 e^4 m}{2\hbar^2 n^2} - \left(\frac{Ze^2}{\hbar c}\right)^4 \frac{mc^2}{2n^4} \left(\frac{n}{j+\frac{1}{2}} - \frac{3}{4}\right), \quad (119.2)$$

where $j = l + \frac{1}{2}$ is the eigenvalue of the total angular momentum operator; the other quantities have the same meaning as in formula (38.17). The energy levels depend now not only on n but also on j . For convenience of comparison with non-relativistic results the formula (119.2) was obtained from the accurate formula by expansion in powers of $Ze^2/\hbar c$.

Accidental degeneracy (see §38) is removed, and energy levels with one and the same value of n but with different j have different values. However, this splitting of levels is very small in comparison with the spacing between neighbouring levels with different n .

* See L. Schiff, *Quantum mechanics* (McGraw-Hill Book Company, New York, 1949).

The degeneracy of states with the same value of j is conserved. For example, for $n = 2$ there are the following three states: $2S_{\frac{1}{2}}$, $2P_{\frac{1}{2}}$ and $2P_{\frac{3}{2}}$. The first two states are degenerate, since they correspond to $n = 2$ and $j = \frac{1}{2}$.

Up to relatively recent times it was assumed that Dirac's theory gave the fine structure of hydrogen levels with a very high degree of accuracy. The distribution of terms, the selection rules and the intensities of lines given by the theory were exactly the same as those found experimentally. It was only in 1953 that Lamb, using radiospectroscopic methods for the measurement, discovered that the $2S_{\frac{1}{2}}$ and $2P_{\frac{1}{2}}$ levels have slightly different energies.

This disagreement between formula (119.2) obtained from the Dirac theory and experiment is associated with a fundamental property of matter, the reality of the vacuum, and in the end not only does it not contradict Dirac's theory but is one of its most brilliant confirmations. New mathematical methods by means of which the Lamb shift was found from the Dirac theory will be described in ch. 14.

§120. The invariance of the Dirac equation with respect to reflection, rotation and Lorentz transformation of coordinates

In §113 we have considered some properties of the Dirac equation. Let us now show that this equation satisfies the conditions of invariance with respect to reflection, rotation and Lorentz transformations. The rotation of the spatial system of coordinates and the Lorentz transformation are linear and orthogonal transformations. We can write them in the form

$$x'_\mu = a_{\mu\nu} x_\nu, \quad x_\nu = a_{\mu\nu} x'_\mu, \quad a_{\mu\nu} a_{\mu\rho} = \delta_{\nu\rho}. \quad (120.1)$$

We find the transformation of the wave function

$$\psi' = S\psi, \quad (120.2)$$

which leaves the Dirac equation invariant under linear transformations (120.1). The transformed wave function satisfies the Dirac equation

$$\gamma_\mu \frac{\partial \psi'}{\partial x'_\mu} + \frac{mc}{\hbar} \psi' = 0. \quad (120.3)$$

The derivatives $\partial/\partial x'_\mu$ can be transformed by means of the relation

$$\frac{\partial}{\partial x'_\mu} = \frac{\partial}{\partial x_\nu} \frac{\partial x_\nu}{\partial x'_\mu} = a_{\mu\nu} \frac{\partial}{\partial x_\nu}. \quad (120.4)$$

Making use of (120.4) we transform eq. (120.3) into the form

$$a_{\mu\nu}\gamma_\mu S \frac{\partial\psi}{\partial x_\nu} + \frac{mc}{\hbar} S\psi = 0. \quad (120.5)$$

If there exists a matrix S^{-1} for which the conditions

$$S^{-1}a_{\mu\nu}\gamma_\mu S = \gamma_\nu \quad \text{or} \quad S^{-1}\gamma_\mu S = \sum_\nu a_{\mu\nu}\gamma_\nu, \quad S^{-1}S = 1 \quad (120.6)$$

are fulfilled, then, multiplying eq. (120.5) by the matrix S^{-1} , we arrive at eq. (113.16).

Let us now find the explicit form of the matrix of the linear transformation S for rotation of the spatial system of coordinates and Lorentz transformations. In the case of the rotation of the system of coordinates in the (x_1, x_2) -plane the coefficients $a_{\mu\nu}$ are defined by the relations

$$\begin{aligned} x'_1 &= x_1 \cos \varphi + x_2 \sin \varphi, \\ x'_2 &= -x_1 \sin \varphi + x_2 \cos \varphi. \end{aligned} \quad (120.7)$$

We shall now show that if the matrix S is chosen in the form

$$S = e^{\frac{i}{2}\varphi\gamma_1\gamma_2}, \quad (120.8)$$

then relations (120.6) are fulfilled. For this we expand the exponential in a series

$$S = 1 + \frac{\varphi}{2}\gamma_1\gamma_2 + \frac{\varphi^2}{2!4}(\gamma_1\gamma_2)^2 + \frac{\varphi^3}{3!8}(\gamma_1\gamma_2)^3 + \frac{\varphi^4}{4!16}(\gamma_1\gamma_2)^4 + \dots$$

Further, making use of the expressions

$$\begin{aligned} (\gamma_1\gamma_2)^2 &= \gamma_1\gamma_2\gamma_1\gamma_2 = -\gamma_1\gamma_1\gamma_2\gamma_2 = -1, \\ (\gamma_1\gamma_2)^3 &= (\gamma_1\gamma_2)^2\gamma_1\gamma_2 = -\gamma_1\gamma_2, \\ (\gamma_1\gamma_2)^4 &= (\gamma_1\gamma_2)^2(\gamma_1\gamma_2)^2 = 1, \end{aligned}$$

we find

$$S = \left(1 - \frac{\varphi^2}{2!4} + \frac{\varphi^4}{4!16} - \dots\right) + \gamma_1\gamma_2 \left(\frac{\varphi}{2} - \frac{\varphi^3}{3!8} + \frac{\varphi^5}{5!32} - \dots\right). \quad (120.9)$$

It is easily seen that the matrix S is equal to

$$S = \cos \frac{1}{2}\varphi + \gamma_1\gamma_2 \sin \frac{1}{2}\varphi. \quad (120.10)$$

Let us now check relations (120.6). The equality $S^{-1}S = 1$ is obvious. Let us find, for example, the expression

$$S^{-1}\gamma_1 S = (\cos \tfrac{1}{2}\varphi - \gamma_1 \gamma_2 \sin \tfrac{1}{2}\varphi) \gamma_1 (\cos \tfrac{1}{2}\varphi + \gamma_1 \gamma_2 \sin \tfrac{1}{2}\varphi).$$

Using the properties of the γ -matrices and elementary trigonometric formulae, we find

$$S^{-1}\gamma_1 S = \gamma_1 \cos \varphi + \gamma_2 \sin \varphi = a_{11}\gamma_1 + a_{12}\gamma_2 = a_{1\nu}\gamma_\nu,$$

which is in complete agreement with (120.6).

We now turn to Lorentz transformations. According to § 10 of Part II, the Lorentz transformation can be treated as a rotation through an imaginary angle $\varphi = i\chi$ in the $(x_1 x_4)$ -plane:

$$x'_1 = x_1 \cosh \chi - x_0 \sinh \chi, \quad \tanh \chi = v/c;$$

$$x'_0 = -x_1 \sinh \chi + x_0 \cosh \chi, \quad \sinh \chi = \frac{v}{c(1-v^2/c^2)^{\frac{1}{2}}}.$$

The matrix S can be found in analogy to (120.8), replacing the angle φ by $i\chi$. Then S assumes the form

$$S = e^{\frac{1}{2}i\chi\gamma_1\gamma_4} = \cosh \tfrac{1}{2}\chi + i\gamma_1\gamma_4 \sinh \tfrac{1}{2}\chi. \quad (120.11)$$

Besides the rotation transformation and the Lorentz transformation it is necessary to consider the transformation of inversion in the origin. Under the inversion of coordinates the spatial coordinates change according to the formulae

$$x_1 \rightarrow -x'_1, \quad x_2 \rightarrow -x'_2, \quad x_3 \rightarrow -x'_3, \quad x_4 = x'_4. \quad (120.12)$$

We have to require that the equation

$$\sum_{i=1}^3 \gamma_i \frac{\partial \psi'}{\partial x'_i} + \gamma_4 \frac{\partial \psi'}{\partial x'_4} + \frac{mc}{\hbar} \psi' = 0$$

remain invariant under the replacement (120.12), and that the wave function undergo the transformation $\psi' = \hat{I}\psi$. It is easily seen that the requirement of invariance will be fulfilled if the operator \hat{I} is of the form

$$\hat{I} = a\beta, \quad (120.13)$$

where a is a certain number. Indeed, making use of (120.12) and (120.13) we obtain

$$-\sum_{i=1}^3 \gamma_i \hat{I} \frac{\partial \psi}{\partial x_i} + \hat{I} \gamma_4 \frac{\partial \psi}{\partial x_4} + \frac{mc}{\hbar} \hat{I} \psi = 0 .$$

Multiplying this equation from the left by β and dividing it by a , we obtain eq. (113.16).

A double inversion transformation brings the system back into its initial state i.e. it corresponds to rotation over the angle 2π . In the last case ψ may change sign. Hence we find the condition imposed upon the quantity

$$a^2 = \pm 1 . \quad (120.14)$$

For what follows we shall need the laws of transformation of the function

$$\bar{\psi} = \psi^\dagger \gamma_4 . \quad (120.15)$$

They can be introduced if it is noted that the function ψ^\dagger satisfies the equation

$$\sum_{i=1}^3 \frac{\partial \psi^\dagger}{\partial x_i} \gamma_i + \frac{mc}{\hbar} \psi^\dagger - \frac{\partial \psi^\dagger}{ic \partial t} \gamma_4 = 0 . \quad (120.16)$$

The requirements of invariance of this equation with respect to the rotation of the spatial system of coordinates and the Lorentz transformation lead to the condition

$$\bar{\psi}' = \bar{\psi} S^{-1} . \quad (120.17)$$

Under the inversion transformation we find $\bar{\psi}' = a^* \bar{\psi} \gamma_4$.

§121. The laws of transformation of bilinear combinations made up of wave functions

Later, in discussing one of the basic problems of modern physics, the problem of the interactions between elementary particle, we shall have to make use of certain properties of bilinear combinations made up of the wave function ψ and the function $\bar{\psi}$ conjugate to it.

As we shall see in what follows, it is necessary for a relativistically invariant formulation of the laws of interaction of nuclear particles to know the laws of transformation of the bilinear combinations of the quantities mentioned under the Lorentz transformation, spatial rotation and inversion. A simple calculation shows that from the components of the wave function and the

γ -matrices one can construct certain bilinear combinations which possess the following transformation properties:

$\bar{\psi}\psi$	one component (scalar) ,
$\bar{\psi}\gamma_i\psi$	four components (4-vector) ,
$\bar{\psi}\gamma_i\gamma_5\psi$	four components (pseudovector) ,
$\bar{\psi}\gamma_i\gamma_k\psi$	six components $i \neq k$ (4-tensor of the second rank) ,
$\bar{\psi}\gamma_5\psi$	one component (pseudoscalar) .

Here the following notation is introduced:

$$\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4 = - \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \quad (121.1)$$

The quantity γ_5 has the following properties:

$$\gamma_5^2 = 1, \quad \gamma_5\gamma_\mu + \gamma_\mu\gamma_5 = 0, \quad \mu = 1, 2, 3, 4.$$

The validity of these relations is easily seen by direct check. We turn to the proof of these transformation properties. When a Lorentz transformation and a spatial rotation are made, one can write by virtue of (120.2) and (120.17)

$$\bar{\psi}'\psi' = \bar{\psi}S^{-1}S\psi = \bar{\psi}\psi.$$

Because $\bar{\psi}'\psi' = a^*\bar{\psi}\gamma_4 a\gamma_4\psi = \bar{\psi}\psi$ the quantity $\bar{\psi}\psi$ also remains invariant under reflection of the system of coordinates. Thus we see that the quantity $\bar{\psi}\psi$ is invariant with respect to an orthogonal transformation.

Further, we shall show that the four quantities $\bar{\psi}\gamma_i\psi$ transform as the components of a four-dimensional vector. When the rotation of the system of coordinates and a Lorentz transformation are carried out we can write

$$\bar{\psi}'\gamma_i\psi' = \bar{\psi}S^{-1}\gamma_i S\psi.$$

In correspondence with formula (120.6) we find

$$\bar{\psi}'\gamma_i\psi' = a_{i\nu}\bar{\psi}\gamma_\nu\psi. \quad (121.2)$$

When inversion of coordinates is carried out we obtain

$$\bar{\psi}'\gamma_i\psi' = a^*a\bar{\psi}\gamma_4\gamma_i\gamma_4\psi = -\bar{\psi}\gamma_i\psi \quad (i \neq 4). \quad (121.3)$$

Thus the quantity $\bar{\psi}\gamma_i\psi$ changes sign under the inversion. Formulae (121.2)

and (121.3) show that the four components indeed form a four-dimensional vector.

We shall now show that the quantity $\bar{\psi}\gamma_5\psi$ represents a pseudoscalar. Under inversion of coordinates we have

$$\bar{\psi}'\gamma_5\psi' = a^* \bar{\psi} \gamma_4 \gamma_1 \gamma_2 \gamma_3 \gamma_4 \gamma_4 \psi.$$

Using the form of the matrix γ_5 and the commutation property of the matrices $\gamma_i\gamma_k + \gamma_k\gamma_i = 2\delta_{ik}$, we easily find

$$\bar{\psi}'\gamma_5\psi' = -\bar{\psi}\gamma_5\psi,$$

which proves the statement which we have already formulated. The quantities $\bar{\psi}\gamma_i\gamma_5\psi$ do not change sign under reflection, and transform as the components of a vector under the rotation and Lorentz transformations. Consequently, we can state that these quantities are the components of a four-dimensional axial vector or pseudovector.

We can convince ourselves of the tensor character of the quantities $\bar{\psi}\gamma_i\gamma_k\psi$ in an analogous way:

$$\bar{\psi}'\gamma_i\gamma_k\psi' = \bar{\psi}S^{-1}\gamma_i\gamma_kS\psi = \bar{\psi}S^{-1}\gamma_iSS^{-1}\gamma_kS\psi = a_{il}a_{km}\bar{\psi}\gamma_l\gamma_m\psi,$$

which is the same as the definition of a tensor.

§ 122. The concept of weak interactions. Parity non-conservation

We have already seen that in addition to the electromagnetic interaction there is also another form of interaction; the strong interaction between nucleons.

It turns out that in addition to the strong interaction there is one more form of interaction which is also of non-electromagnetic character and is called the weak interaction (see below, § 130).

Weak interactions, which cannot bind nucleons in the nucleus, play an important role in the physics of elementary and nuclear particles. They are responsible for the radioactive decay of nuclei with the emission of light particles, electrons and neutrinos. In other words, the weak interaction between elementary particles leads to β -decay.

The theory of weak interactions has recently achieved considerably successes. However, a consideration of the relevant problems is possible only within the framework of the quantum field theory and hence we shall confine ourselves only to some comments. First of all we note that the Dirac equation (113.7) can be considered as the equation for a certain electron—positron

field ψ . We have already mentioned such a field approach in §112 where we considered the Klein—Gordon—Fock equation. In the field description particles are considered as the excitation quanta of the corresponding field (for example, photons are the excitation quanta of the electromagnetic field (see §101 and §102)). Then the function ψ should be considered as an operator in the space of occupation numbers (see formula (99.26) of the theory of second quantization). Of course, passing to the 'field' description we give up the one-particle interpretation of the Dirac equation. The operator ψ has non-zero matrix elements corresponding to the absorption of an electron and the production of a positron, whereas the operator ψ^\dagger has non-zero matrix elements corresponding to the production of an electron and the absorption of a positron. Such considerations are general and apply also to other particles (μ -mesons, neutrinos, nucleons and so on).

Let us now consider any process, for example the decay of a μ -meson, with the emission of a neutrino and an antineutrino

$$\mu \rightarrow e + \nu + \bar{\nu}.$$

We recall that by definition the neutrino is understood to be the particle emitted in the positron decay of the proton

$$p \rightarrow e^+ + n + \nu,$$

and the antineutrino the particle emitted in the β -decay of the neutron

$$n \rightarrow p + e^- + \bar{\nu}.$$

Experimental data available at present show that these particles are not identical.

The process of decay of the μ -meson involves four particles with spin $\frac{1}{2}$, four fermions.

For the description of the μ -meson, the electron, and the neutrino we introduce respectively the operators ψ_μ , ψ_e and ψ_ν each of which satisfies the corresponding Dirac equation. The basic problem consists now in choosing the interaction leading to the decay. For this it is necessary to formulate the interaction Hamiltonian

$$\hat{H}' = \int \hat{H}' dV, \quad (122.1)$$

where \hat{H}' is the density of the interaction Hamiltonian.

Since the ψ'_S are operators in the space of occupation numbers, the density of the interaction Hamiltonian must contain these operators, as in non-relativistic physics (§99).

From the structure of expression (122.1) it is seen that the density of the

interaction Hamiltonian \hat{H}' (we shall sometimes omit the word 'density') must be a relativistic scalar (invariant with respect to the rotation and Lorentz transformations). Until the mid-fifties there were no doubts as to the existence of symmetry with respect to 'the right' and to 'the left', i.e. it was assumed that the parity conservation law holds for all interactions. Hence it was assumed that the density \hat{H}' must also be invariant with respect to the inversion transformation. The requirement of relativistic invariance strongly restricts the class of possible expressions for \hat{H}' . Namely, since in the theory of relativity any interaction has the character of a short-range action, the values of the characteristics of all particles (the operators ψ) must be taken at one point of space and at one instant of time.

For the process of β -decay involving four fermions

$$A + B \rightarrow C + D. \quad (122.2)$$

Fermi proposed the simplest law of interaction in the form

$$\hat{H}' \sim (\bar{\psi}_C \Gamma \psi_A) (\bar{\psi}_D \Gamma \psi_B) + \text{Herm.conj.}, \quad (122.3)$$

where Herm.conj. denotes the Hermitian conjugate expression, and the value of all operators ψ_i is taken at one point. The quantity Γ can have the following forms:

$\Gamma_1 = 1$	scalar covariant,
$\Gamma_2 = \gamma_\mu$	vector covariant,
$\Gamma_3 = \sigma_{\mu\nu}$	tensor covariant,
$\Gamma_4 = \gamma_\mu \gamma_5$	pseudovector covariant,
$\Gamma_5 = \gamma_5$	pseudoscalar covariant,

where $\sigma_{\mu\nu} = -\frac{1}{2}i(\gamma_\mu \gamma_\nu - \gamma_\nu \gamma_\mu)$, $\mu, \nu = 1, 2, 3, 4$, and the summation in (122.3) is carried out from one to four over repeated vector indices.

The Hamiltonian density (122.3) does not involve the derivatives of the operators ψ and $\bar{\psi}$. This form of the interaction Hamiltonian is called 'coupling without derivatives'. We shall come back to the problem of the absence of derivatives in the law of interaction below.

We have, in §121, established the transformation properties of bilinear combinations of the type $(\bar{\psi}_C \Gamma \psi_A)$. Since the operator \hat{H}' contains products in which the quantities Γ are involved twice, it is a scalar for all Γ . Thus, for example, for the vector covariant of the interaction we have

$$\hat{H}' = g_2 (\bar{\psi}_C \gamma_\mu \psi_A) (\bar{\psi}_D \gamma_\mu \psi_B) + \text{Herm.conj.},$$

where the constant g_2 is called the coupling constant or the interaction con-

stant of the vector covariant. The vector covariant is constructed as the scalar product of two four-dimensional vectors (the summation from 1 to 4 is carried out over μ). The addition of the Hermitian conjugate terms makes the operator Hermitian.

In the general case the Hamiltonian density represents the sum of all five types of interaction. The expression written satisfies the requirements of the theory of relativity and, besides the characteristics of the particles, the operators ψ , contains only the interaction constant and the matrices involved in the Dirac equation.

We now come back to the process of decay of the μ -meson with the emission of a neutrino and an antineutrino. Since the operator ψ_ν describes the emission of an antineutrino as well as the absorption of a neutrino, the process of decay of the μ -meson is equivalent to a process with the absorption of a neutrino

$$\mu + \nu \rightarrow e + \nu.$$

Correspondingly \hat{H}' is of the form

$$\hat{H}' = \sum_{k=1}^5 g_k (\bar{\psi}_e \Gamma_k \psi_\nu) (\bar{\psi}_\nu \Gamma_k \psi_\mu) + \text{Herm.conj.} \quad (122.4)$$

The use of the Hamiltonian (122.3) led to some success in the construction of a theory of β -decay. As will be clear from what follows, the Hamiltonian (122.4) became a basis for working out the modern theory of β -decay.

The further considerable development of the theory was associated with the discovery of parity non-conservation in weak interactions. The assumption of parity non-conservation in weak interactions was made by Lee and Yang* on the basis of available data on two types of K-meson decay.

K-mesons represent a group of elementary particles (a positive, a negative and two neutral ones) having zero spin and a mass of about 966 electron masses. All K-mesons are unstable and decay with a lifetime of 1.2×10^{-8} sec for the charged mesons and 10^{-10} sec and 6×10^{-8} sec for the two neutral mesons. It turns out that in addition to the decay into a μ -meson and a neutrino, K-mesons can decay according to the schemes

* T.D.Lee and C.N.Yang, *New properties of symmetry of elementary particles*, Phys. Rev. 102 (1956) 290; 104 (1956) 254.

$$K^+ \rightarrow \pi^+ + \pi^0, \quad (\theta\text{-decay}),$$

$$K^+ \rightarrow \begin{cases} \pi^+ + \pi^- + \pi^+, \\ \pi^+ + \pi^0 + \pi^0 \end{cases} \quad (\tau\text{-decay}).$$

The possibility of decay of the K-meson into two or three π -mesons directly contradicts the parity conservation law. Indeed, the analysis of the properties of π -mesons and of their angular distribution shows that the parity of a system of two mesons differs from that of a system of three mesons.

Still more definite indications of parity non-conservation were obtained subsequently in studying the β -decay of polarized nuclei of ^{60}Co . The nuclei of ^{60}Co have a spin σ different from zero. This fact imposes certain requirements upon the angular distribution of the β -electrons emitted by them. Namely, it follows from the parity conservation law that the distribution of the electrons must possess symmetry with respect to the direction of the vector σ . The number of electrons emerging at angles θ and $180^\circ - \theta$ with respect to the direction of σ must be the same. Indeed, if the number of electrons entering solid angle $d\Omega$ is written in the form

$$dI = F(\theta) d\Omega,$$

where F is a certain function of the angle θ between the vectors \mathbf{p} and σ , then this relation should not be violated under the inversion transformation.

The vector σ , being an axial vector, does not change, whereas the polar vector \mathbf{p} changes sign under the inversion transformation. Hence the angle θ transforms under inversion: $\theta \rightarrow 180^\circ - \theta$. Thus the parity conservation law requires the invariance of the distribution function

$$F(\theta) = F(180^\circ - \theta).$$

Direct measurements showed that the angular distribution of the β -electrons emitted by polarized nuclei of ^{60}Co does not possess the symmetry mentioned. On the contrary, the electrons emerge preferentially in the direction opposite to the orientation of the spin of the nucleus. Thus the β -decay of polarized nuclei demonstrates directly the violation of the parity conservation law.

The parity conservation law, as we have seen in §33, is associated with symmetry properties of space. Violation of parity would mean that space possesses no symmetry and that the notions of 'right' and 'left' in it are of absolute character. Such an interpretation would lead to extremely grave difficulties in interpreting all the laws of physics. It appeared to be com-

pletely incomprehensible how space, could be asymmetric while remaining homogeneous and isotropic.

A way out of this difficulty was proposed by Landau. According to Landau's hypothesis, the particles themselves are asymmetric, not space. Landau* proposed the principle of combined parity, according to which all physical laws must remain invariant under combined inversion, space inversion and the simultaneous replacement of particles by antiparticles (so-called charge conjugation). As examples of the latter we can mention the replacement of electrons by positrons, protons by antiprotons and so on.

Parity non-conservation in weak interactions leads to the fact that the Hamiltonian \hat{H}' must no longer necessarily be a scalar with respect to reflection. Consequently, in the general case the Hamiltonian (122.3) must be supplemented, by introducing into it terms which change sign under the reflection of coordinates

$$\hat{H}' = \sum_{k=1}^5 [g_k(\bar{\psi}_C \Gamma_k \psi_A)(\bar{\psi}_D \Gamma_k \psi_B) + g'_k(\bar{\psi}_C \Gamma_k \psi_A)(\bar{\psi}_D \Gamma_k \gamma_5 \psi_B)] + \text{Herm.conj.} \quad (122.5)$$

The second component of each term of the sum is a pseudoscalar. The constants g'_k , generally speaking, are not the same as the constants g_k . One would think that the increase in the number of constants makes the interpretation of available experimental data and its comparison with conclusions from theory difficult. However, as a matter of fact, parity non-conservation opened new possibilities and led to the formulation of the universal law of the four-fermion interaction.

§123. Two-component neutrino theory. The universal four-fermion interaction

The discovery of parity non-conservation in weak interactions made it possible to formulate the theory of the longitudinal or two-component neutrino**. The theory of the two-component neutrino is based on the assumption that the mass of the neutrino is not simply small but exactly equal to zero. Since the neutrino has spin one-half, it is described by the Dirac

* L.D.Landau, Soviet Physics JETP 5 (1957) 336.

** L.D.Landau, Soviet Physics JETP 5 (1957) 336; Nuclear Physics 3 (1957) 127; A.Salam, Nuovo Cimento 5 (1957) 299.

equation, which for $m = 0$ for states with given momentum \mathbf{p} is of the form (see (115.3))

$$Eu = (\boldsymbol{\alpha} \cdot \mathbf{p})u, \quad E = \pm |\mathbf{p}|. \quad (123.1)$$

(In this section and further on we use the system of units in which $\hbar = 1$ and $c = 1$.) We can pass from eq. (123.1) for the four-component function u to the equation for two-component functions. Setting $u = \frac{1}{2}\sqrt{2} \begin{pmatrix} w \\ w' \end{pmatrix}$ and taking into account that $\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}$, we rewrite eq. (123.1) in the form

$$Ew = (\boldsymbol{\sigma} \cdot \mathbf{p})w', \quad Ew' = (\boldsymbol{\sigma} \cdot \mathbf{p})w. \quad (123.2)$$

Adding up and subtracting eq. (123.2) we obtain

$$E\varphi_+ = (\boldsymbol{\sigma} \cdot \mathbf{p})\varphi_+, \quad E\varphi_- = -(\boldsymbol{\sigma} \cdot \mathbf{p})\varphi_-, \quad (123.3)$$

where

$$\varphi_+ = \frac{1}{\sqrt{2}}(w + w'), \quad \varphi_- = \frac{1}{\sqrt{2}}(w - w').$$

We see that the two-component functions φ_+ and φ_- satisfy equations of the first order. Of course, if parity were conserved we could not make use of the superposition of the functions w and w' , since these functions transform differently under the inversion of the system of coordinates. Indeed, since $\boldsymbol{\sigma}$ is an axial vector and \mathbf{p} is a polar vector, the product $(\boldsymbol{\sigma} \cdot \mathbf{p})$ is a pseudoscalar. Then from (123.2) we see that if w transforms under the reflection as a polar spinor, w' transforms as a pseudospinor, and vice versa.

Choosing the direction of the vector \mathbf{p} to be the z -axis, we obtain from (123.3)

$$\sigma_z \varphi_+ = \varphi_+ \quad \text{for} \quad E = |\mathbf{p}|, \quad (123.4)$$

$$\sigma_z \varphi_+ = -\varphi_+ \quad \text{for} \quad E = -|\mathbf{p}|$$

and

$$\sigma_z \varphi_- = -\varphi_- \quad \text{for} \quad E = |\mathbf{p}|, \quad (123.5)$$

$$\sigma_z \varphi_- = \varphi_- \quad \text{for} \quad E = -|\mathbf{p}|.$$

We see that the functions φ_+ and φ_- describe states whose polarization (the spin component along the z -axis) is unambiguously related to the sign of the energy. Thus the function φ_- describes the state polarized against the direction of the momentum for $E = |\mathbf{p}|$ and along the direction of the momentum for $E = -|\mathbf{p}|$, whereas the function φ_+ describes the state polarized along the direction of the momentum for $E = |\mathbf{p}|$ and against the direction of the momentum for $E = -|\mathbf{p}|$.

In the theory of the two-component neutrino it is assumed that the neutrino ($E=|\mathbf{p}|$) and antineutrino ($E=-|\mathbf{p}|$) are described by the function φ_- , i.e. that the neutrino is always polarized against the direction of the momentum, and the antineutrino always along the direction of the momentum. Of course, it might equally well be assumed that the neutrino and antineutrino are described by the function φ_+ , but this would lead to conclusions which are in disagreement with experimental data. If the energy of the antineutrino is also assumed to be positive $E=|\mathbf{p}|$, then the antineutrino will be described by the function φ_+ (see the first equation of (123.4) and the second equation of (123.5)).

Under the inversion of the system of coordinates the axial vector $\boldsymbol{\sigma}$ does not change whereas the polar vector \mathbf{p} reverses its direction. Consequently, the neutrino then goes over into the antineutrino and vice versa, in accordance with the ideas put forward by Landau (combined parity conservation).

The entire conclusion presented is based on the assumption that the mass of the neutrino is exactly equal to zero. This also follows immediately from the following obvious considerations. If the mass of the neutrino were not equal to zero, then it would move with a velocity less than the velocity of light. There would then exist an inertial system of coordinates, moving with respect to the laboratory system of coordinates, with a velocity larger than the velocity of the neutrino, in which the direction of the momentum of the neutrino would be reversed. Since the direction of the spin does not change under such a transformation, we would have in one inertial system of coordinates a neutrino, and in the other an antineutrino, i.e. we would arrive at a contradiction, since the neutrino and antineutrino by assumption are not identical.

The theory of the two-component neutrino can easily be formulated within the framework of the usual mathematical apparatus, i.e. by means of four-component functions. Namely, it is easily seen (see (113.16)) that if the bispinor ψ is the solution of the Dirac equation with a rest mass equal to zero

$$\gamma_\mu \frac{\partial \psi}{\partial x_\mu} = 0,$$

then the functions ψ_+ and ψ_- will also be solutions of this equation

$$\psi_+ = \frac{1}{\sqrt{2}} (1 - \gamma_5) \psi, \quad \psi_- = \frac{1}{\sqrt{2}} (1 + \gamma_5) \psi \quad (123.6)$$

or respectively for states with a definite momentum

$$u_+ = \frac{1}{\sqrt{2}} (1 - \gamma_5) u, \quad u_- = \frac{1}{\sqrt{2}} (1 + \gamma_5) u, \quad (123.7)$$

where u satisfies eq. (123.1).

It is easily seen that the functions u_+ and u_- are expressed in terms of φ_+ and φ_- . Indeed, setting $u = \frac{1}{2}\sqrt{2} \begin{pmatrix} w \\ w' \end{pmatrix}$ and taking into account that $\gamma_5 = -\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, where the four-by-four matrix is written in terms of two-by-two matrices, we obtain

$$\begin{aligned} u_+ &= \frac{1}{2} \begin{pmatrix} w + w' \\ w + w' \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_+ \\ \varphi_+ \end{pmatrix}, \\ u_- &= \frac{1}{2} \begin{pmatrix} w - w' \\ -(w - w') \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_- \\ -\varphi_- \end{pmatrix}. \end{aligned} \quad (123.8)$$

The function ψ_-

$$\psi_- = u_- e^{i(\mathbf{p} \cdot \mathbf{r} - Et)}, \quad (123.9)$$

describes the neutrino for $E = |\mathbf{p}|$. It is the eigenfunction of the spin component operator \hat{s}_z ,

$$\hat{s}_z = \begin{pmatrix} \sigma_z & 0 \\ 0 & \sigma_z \end{pmatrix},$$

corresponding to the eigenvalue -1 ,

$$\hat{s}_z u_- = -u_- \quad \text{for} \quad E = |\mathbf{p}| \quad (\text{neutrino}).$$

The antineutrino ($E = |\mathbf{p}|$) is described by the function u_+ and ψ_+ respectively. Under the action of the operator \hat{s}_z we have

$$\hat{s}_z u_+ = u_+ \quad \text{for} \quad E = |\mathbf{p}| \quad (\text{antineutrino}).$$

We also note that the functions ψ_+ and ψ_- are eigenfunctions of the operator γ_5 . Indeed, since $\gamma_5^2 = 1$, it follows from (123.6) that

$$\gamma_5 \psi_+ = -\psi_+, \quad \gamma_5 \psi_- = \psi_-. \quad (123.10)$$

The operator γ_5 is called the helicity operator. To the eigenvalue $\gamma_5 = +1$ there corresponds left-handed helicity, while to the eigenvalue $\gamma_5 = -1$ there corresponds right-handed helicity. The above results can be given an obvious interpretation in terms of the helicity operator: there is a strong correlation between the direction of the momentum vector and the direction of the spin vector of a particle. For the neutrino the spin σ is antiparallel to \mathbf{p} ($\gamma_5 = -1$),

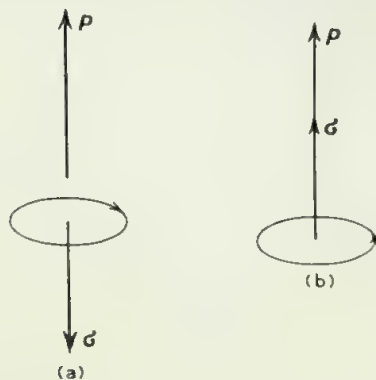


Fig. V.29

whereas for the antineutrino it is parallel ($\gamma_5 = -1$). If the spin is represented in an obvious way as the rotation of the particle, then the neutrino rotates as a left-handed helix about the axis \mathbf{p} (fig. V.29). Under the space inversion the direction of \mathbf{p} reverses, whereas the vector $\boldsymbol{\sigma}$ remains unchanged. There is no neutrino having a spin of 'irregular' orientation. Hence under the reflection of space coordinates it is necessary to allow for the transformation of a neutrino into an antineutrino, in correspondence with the principle of combined parity.

Gell-Mann and Feynman* put forward the hypothesis that the property of helicity is of a general character and is a characteristic of all fermions and not only of the neutrino. According to this hypothesis the transformation (123.6) should hold for all four fermions involved in the process of weak interaction (122.2). This means that in the general expression (122.5) the operators ψ_C , ψ_D , ψ_A and ψ_B should be replaced respectively by the operators

$$\begin{aligned} \chi_C &= \frac{1}{\sqrt{2}} (1 + \gamma_5) \psi_C, & \chi_A &= \frac{1}{\sqrt{2}} (1 + \gamma_5) \psi_A, \\ \chi_D &= \frac{1}{\sqrt{2}} (1 + \gamma_5) \psi_D, & \chi_B &= \frac{1}{\sqrt{2}} (1 + \gamma_5) \psi_B. \end{aligned}$$

Let us elucidate this assumption. We note, first of all, that the operators χ

* R. Feynman and M. Gell-Mann, Phys. Rev. 109 (1958) 193.

are actually two-component operators. They are expressed analogously to (123.8) in terms of the spinors w and w' involved in

$$\chi = \frac{1}{2} \begin{pmatrix} w - w' \\ -(w + w') \end{pmatrix}.$$

Since $\chi = \frac{1}{2}\sqrt{2}(1+\gamma_5)\psi$, then, acting on both sides of the equation with the operator

$$\frac{1}{\sqrt{2m}} \left[\gamma_\mu \left(\frac{\partial}{\partial x_\mu} - ieA_\mu \right) - m \right]$$

and taking into account the Dirac equation

$$\left[\gamma_\mu \left(\frac{\partial}{\partial x_\mu} - ieA_\mu \right) + m \right] \psi = 0,$$

we obtain

$$-\frac{1}{\sqrt{2m}} \left[\gamma_\mu \left(\frac{\partial}{\partial x_\mu} - ieA_\mu \right) - m \right] \chi = \psi. \quad (123.11)$$

Substituting ψ in the form of (123.11) into the Dirac equation we obtain the equation of second order satisfies by the operator χ

$$\left[\left(\frac{\partial}{\partial x_\mu} - ieA_\mu \right)^2 + e\sigma_{\mu\nu}F_{\mu\nu} - m^2 \right] \chi = 0, \quad (123.12)$$

where

$$\sigma_{\mu\nu} = -\frac{i}{2}(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)$$

and

$$F_{\mu\nu} = \frac{\partial}{\partial x_\mu} A_\nu - \frac{\partial}{\partial x_\nu} A_\mu.$$

Feynman and Gell-Mann assumed that the operator χ is more fundamental than the operator ψ and thus that the interaction Hamiltonian (122.3) should not contain the derivatives of the operator χ . Hence, in view of (123.11) the interaction Hamiltonian should involve the operator χ and not ψ .

Such an assumption leads immediately to the fact that of all the covariants of interaction only the vector covariant and the axial-vector covariant (with the same constants) turn out to be possible, whereas all other covariants give zero.

Let us show, for example, that the pseudoscalar covariant reduces to zero.

$$\hat{H}' = \frac{1}{4}g_5((1+\gamma_5)\bar{\psi}_C\gamma_5(1+\gamma_5)\psi_A)((1+\gamma_5)\bar{\psi}_D\gamma_5(1+\gamma_5)\psi_B) + \\ + \frac{1}{4}g_5'((1+\gamma_5)\bar{\psi}_C\gamma_5(1+\gamma_5)\psi_A)((1+\gamma_5)\bar{\psi}_D(1+\gamma_5)\psi_B) + \text{Herm.conj.} .$$

But

$$\gamma_5(1+\gamma_5) = 1 + \gamma_5 .$$

Consequently the two terms are identical. Further,

$$\overline{(1+\gamma_5)\psi} = ((1+\gamma_5)\psi)^\dagger \gamma_4 = \psi^\dagger (1+\gamma_5) \gamma_4 = \bar{\psi}(1-\gamma_5) .$$

Hence the following terms appear in the parentheses:

$$(1-\gamma_5)(1+\gamma_5) = 1 - \gamma_5^2 = 0 .$$

The scalar and tensor covariants also give zero.

Taking into account that

$$\gamma_5(1+\gamma_5) = (1+\gamma_5) , \quad (1+\gamma_5)^2 = 2(1+\gamma_5) ,$$

we have for the vector and axial-vector covariants

$$\begin{aligned} \hat{H}' &= \frac{1}{4}(g_2+g_2')[\bar{\psi}_C(1-\gamma_5)\gamma_\nu(1+\gamma_5)\psi_A][\bar{\psi}_D(1-\gamma_5)\gamma_\nu(1+\gamma_5)\psi_B] + \\ &+ \frac{1}{4}(g_4+g_4')[\bar{\psi}_C(1-\gamma_5)\gamma_\nu\gamma_5(1+\gamma_5)\psi_A] \times \\ &\times [\bar{\psi}_D(1-\gamma_5)\gamma_\nu\gamma_5(1+\gamma_5)\psi_B] + \text{Herm.conj.} = \\ &= \frac{1}{4}(g_2+g_2'+g_4+g_4') \times \\ &\times [\bar{\psi}_C\gamma_\nu 2(1+\gamma_5)\psi_A][\bar{\psi}_D\gamma_\nu 2(1+\gamma_5)\psi_B] + \text{Herm.conj.} \equiv \\ &\equiv f(\bar{\psi}_C\gamma_\nu(1+\gamma_5)\psi_A)(\bar{\psi}_D\gamma_\nu(1+\gamma_5)\psi_B) + \text{Herm.conj.} . \end{aligned} \quad (123.13)$$

On the basis of a careful analysis of experimental data Sudarshan and Marshak arrived at exactly the same form of the Hamiltonian for the four-fermion interaction.

The interaction Hamiltonian (123.13) gives the universal law of the four-fermion interaction with only one coupling constant f . In analyzing concrete processes by means of the Hamiltonian (123.13) it is also necessary to take into account the so-called leptonic charge conservation law. Leptons are light particles taking part in weak interaction processes, namely: electrons e^- , μ^- , mesons and the neutrino ν . The particles e^+ , μ^+ and $\bar{\nu}$ are called antileptons. Leptons are assigned the leptonic charge $+1$, and antileptons -1 . For other particles, for example nucleons, the leptonic charge is assumed to be equal to zero. The total leptonic charge (the algebraic sum of the leptonic charges) must be conserved in the reaction (see ch. 15).

Let us consider, for example, the decay of the μ^- -meson

$$\mu^- \rightarrow e^- + \nu + \bar{\nu}. \quad (123.14)$$

A decay with the emission of two neutrinos (or antineutrinos) is evidently forbidden by the leptonic charge conservation law.

The interaction Hamiltonian (123.14), in correspondence with (123.13), is of the form

$$\hat{H}' = f(\bar{\psi}_e \gamma_\mu (1 + \gamma_5) \psi_\nu)(\bar{\psi}_\nu \gamma_\mu (1 + \gamma_5) \psi_{\mu-}) + \text{Herm.conj.} \quad (123.15)$$

For the process of β -decay of the neutron we have, correspondingly

$$\hat{H}' = f(\bar{\psi}_e \gamma_\mu (1 + \gamma_5) \psi_\nu)(\bar{\psi}_p \gamma_\mu (1 + \gamma_5) \psi_n) + \text{Herm.conj.} \quad (123.16)$$

Knowing the interaction Hamiltonian it is easy to determine the probability of the corresponding process by ordinary methods of perturbation theory. The universal law of the four-fermion interaction, proposed by Gell-Mann and Feynman and by Sudarshan and Marshak is quantitatively confirmed by a vast amount of experimental data.

Some Problems of Quantum Electrodynamics

§124. The Green's function of the Dirac equation

The theory of the interaction of non-relativistic charged particles with the electromagnetic field, presented in ch. 12, is easily generalized to the case of relativistic particles. However, calculations of higher approximations of perturbation theory (expansion in powers of $e^2/\hbar c$) led to diverging expressions whose physical meaning was not clear. Thus, for example, the intrinsic energy of the electron turned out to be infinite, as in classical electrodynamics. Corrections to the scattering cross sections calculated in the second order approximation of perturbation theory also turned out to be infinitely large, and so on. All this pointed to a limited region of applicability of the mathematical apparatus of quantum electrodynamics. At the same time good agreement with experimental data on cross sections for different processes calculated in the first non-vanishing approximation of perturbation theory indicated the validity of the general ideas and methods of the theory.

An increase in the accuracy of experimental methods of investigation led recently to the establishment of new facts which had no explanation in quantum electrodynamics. Namely, in 1947, in addition to the discovery by Lamb of the shift of the $2^2S_{1/2}$ and $2^2P_{1/2}$ levels of the hydrogen atom which, according to the Dirac theory, should coincide, Rabi established that the value of the magnetic moment of the electron differs somewhat from a Bohr

magneton. The discovery of these phenomena led to a further intense development of quantum electrodynamics. Very important roles in the development of the theory were played by the studies of Bethe, Feynman, Dyson, Schwinger, Tomonaga and others*. In particular, Feynman proposed a new method of calculation which made it possible to simplify considerably all the calculations and also to give them an obvious physical meaning**. Within the framework of this book we can present only the most general outlines of Feynman's method. A detailed exposition of Feynman's method as well as numerous examples of its application to concrete problems can be found in the articles and monographs cited below.

The method of Green's function is the basis of the mathematical apparatus of Feynman's theory. We now turn directly to the exposition of the method proposed by Feynman. First of all we write the Dirac equation in a form which is more compact and convenient for these calculations. For this we introduce the operator $\hat{\nabla}$

$$\hat{\nabla} \equiv \sum_{\mu=1}^4 \gamma_{\mu} \frac{\partial}{\partial x_{\mu}} \equiv \gamma_{\mu} \frac{\partial}{\partial x_{\mu}},$$

where $x_4 = ix_0 = it$. In this notation the Dirac equation has the form†

$$(\hat{\nabla} + m)\psi = 0. \quad (124.1)$$

Analogously to what we did in non-relativistic theory for the Schrödinger equation (see §29) we introduce the Green's function $K(2,1)$ of the Dirac equation (124.1). The Green's function $K(2,1)$ by definition satisfies the equation

$$(\hat{\nabla}_2 + m)K(2,1) = i^{-1}\delta^4(2,1). \quad (124.2)$$

Here and in what follows the numbers 1 and 2 denote the set of four coor-

* A detailed bibliography is given in the book of S.Schweber, H.Bethe and F.de Hoffman, *Mesons and fields* (Row, Peterson and Company, Evanston, Illinois and White Plains, New York, 1956). For a detailed exposition of quantum electrodynamics see also A.I.Akhiezer and V.B.Berestetskii, *Quantum electrodynamics* (Interscience Publ., New York, 1965).

** R.P.Feynman, Phys. Rev. 76 (1949) 749. See also the monographs cited above.

† We use a notation somewhat different from that introduced by Feynman. A similar notation is adopted, for example, in the book of A.I.Akhiezer and V.B.Berestetskii. We also note (as on p. 510) that in this chapter we assume that $\hbar = 1$ and $c = 1$.

dinates x_μ , and $\hat{\nabla}_2$ is the operator acting on the variables $x_{2\mu}$. The symbol $\delta^4(2,1)$ denotes a 4-dimensional δ -function equal to

$$\delta^4(2,1) = \delta^4(x_2 - x_1) = \delta(\mathbf{r}_2 - \mathbf{r}_1) \delta(t_2 - t_1). \quad (124.3)$$

We seek the solution of eq. (124.1) in the momentum representation. In other words, we expand the function $K(2,1)$ in the Fourier integral

$$K(2,1) = \int_{-\infty}^{\infty} S(p) \exp[ip_\mu(x_{2\mu} - x_{1\mu})] d^4p, \quad (124.4)$$

where

$$d^4p = d^3p dp_0 = dp_x dp_y dp_z dp_0 = dp_1 dp_2 dp_3 dp_0.$$

p_μ is the 4-dimensional momentum vector, and $p_4 = ip_0$. The summation from 1 to 4 is carried out over the index μ . In order not to overload the formulae with indices, we shall in what follows omit the index μ if this cannot lead to misunderstanding. Also expanding $\delta^4(2,1)$ in a Fourier integral according to the formula (see Appendix III in Vol. 1)

$$\delta^4(2,1) = \frac{1}{(2\pi)^4} \int_{-\infty}^{\infty} e^{ip(x_2 - x_1)} d^4p \quad (124.5)$$

and substituting expressions (124.4) and (124.5) into (124.2), we find

$$(\hat{\nabla}_2 + m) \int S(p) e^{ip(x_2 - x_1)} d^4p = \frac{1}{(2\pi)^4 i} \int e^{ip(x_2 - x_1)} d^4p. \quad (124.6)$$

The action of the operator $(\hat{\nabla}_2 + m)$ gives

$$(\hat{\nabla}_2 + m) e^{ip(x_2 - x_1)} = (i\hat{p} + m) e^{ip(x_2 - x_1)}, \quad (124.7)$$

where $\hat{p} = p_\mu \gamma_\mu$. (We stress that in this chapter the mark \wedge has a meaning different from that in the preceding chapters of the book.)

Equating the Fourier components in (124.6) and taking into account (124.7), we can write the formal solution for $S(p)$

$$S(p) = \frac{1}{(2\pi)^4} \frac{1}{i} \frac{1}{i\hat{p} + m} = \frac{-1}{(2\pi)^4} \frac{i\hat{p} - m}{(i\hat{p} + m)(i\hat{p} - m)} = \frac{i}{(2\pi)^4} \frac{i\hat{p} - m}{\hat{p}^2 + m^2}, \quad (124.8)$$

where

$$\hat{p}^2 = \sum_{\mu, \nu} p_\mu p_\nu \gamma_\mu \gamma_\nu = \frac{1}{2} \sum_{\mu, \nu} (\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu) p_\mu p_\nu = \sum_{\mu} p_\mu p_\mu = p^2.$$

We have made use of the anticommutativity of the matrices γ . For $K(2, 1)$ we have correspondingly

$$K(2, 1) = \frac{i}{(2\pi)^4} \int \frac{i\hat{p} - m}{p^2 + m^2} e^{ip(x_2 - x_1)} d^4 p. \quad (124.9)$$

This expression is conveniently written in the form

$$K(2, 1) = i(\hat{\nabla}_2 - m)I(2, 1), \quad (124.10)$$

where $I(2, 1)$ is an integral which depends only on ordinary variables but not on the Dirac matrices and is equal to

$$I(2, 1) = \frac{1}{(2\pi)^4} \int \frac{e^{ip(x_2 - x_1)}}{p^2 + m^2} d^4 p. \quad (124.11)$$

We carry out the integration over the variable p_0 :

$$I(2, 1) = -\frac{1}{(2\pi)^4} \int e^{i\mathbf{p} \cdot (\mathbf{r}_2 - \mathbf{r}_1)} d^3 p \int \frac{e^{-ip_0(t_2 - t_1)}}{p_0^2 - E_p^2} dp_0,$$

where $E_p = +(p^2 + m^2)^{\frac{1}{2}}$. If p_0 is considered as a certain complex variable, then the integration in the plane of this complex variable is carried out over the entire real axis. However, the integrand has poles on this axis at the points $p_0 = E_p$ and $p_0 = -E_p$. Consequently, for the integral (124.11) to have a definite meaning it is necessary to define the rule of circumventing these poles. Feynman proposed the following rule: the left pole is circumvented from below, and the right pole from above. To carry this out one has to add to the mass m an infinitesimal negative imaginary part which in the final result should be made to tend to zero: $m \rightarrow m - i\delta$, $\delta > 0$.

As a matter of fact, E_p then also receives an infinitesimal negative imaginary part, and correspondingly the poles of the integrand are situated as shown in fig. V.30. We can now carry out the integration, closing the contour of integration with an infinitely large semicircle and calculating residues at corresponding poles. Since an exponential function stands within the integral sign, the contour of integration is closed below at $t_2 > t_1$ and above at $t_2 < t_1$. Correspondingly for $t_2 > t_1$ the residue is taken at the point $p_0 = E_p$, and for $t_2 < t_1$ at the point $p_0 = -E_p$. Thus we obtain

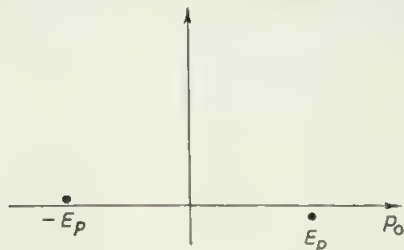


Fig. V.30

$$\begin{aligned}
 I(2,1) &= \frac{i}{16\pi^3} \int \frac{1}{E_p} \exp [ip \cdot (r_2 - r_1) - iE_p(t_2 - t_1)] d^3p \quad (t_2 > t_1), \\
 I(2,1) &= \frac{i}{16\pi^3} \int \frac{1}{E_p} \exp [ip \cdot (r_2 - r_1) + iE_p(t_2 - t_1)] d^3p \quad (t_2 < t_1),
 \end{aligned}
 \tag{124.12}$$

We see that since $E_p > 0$, then for $t_2 > t_1$ only states with a positive energy give a contribution, whereas for $t_2 < t_1$ correspondingly only states with a negative energy contribute. We note that the result obtained differs substantially from that obtained in non-relativistic theory. Indeed, in non-relativistic theory Green's function was assumed (see (29.3)) to be equal to zero for $t_2 < t_1$. We would also obtain an analogous expression in the relativistic case if both poles were circumvented from above, which corresponds to the replacement $p_0 \rightarrow p_0 + i\delta$ in the integrand.

A calculation analogous to that which we have just carried out shows that for such a replacement, there would correspond to a time $t_2 > t_1$ a summation over positive as well as negative energies, and for $t_2 < t_1$ we would obtain $I = 0$. However, from what follows it will be seen that the use of the Green's function proposed by Feynman and defined by formulae (124.12) is much more convenient.

By means of the Green's function introduced it is possible to construct the solution of the Dirac equation, i.e. to obtain a formula analogous to the non-relativistic relation (29.2). For this purpose it is simplest to make use of Gauss' theorem in 4-space ($d^4x = d^3x dt$)

$$i \int \frac{\partial F_\mu(x)}{\partial x_\mu} d^4x = \int_S F_\mu(x) n_\mu d\sigma(x), \tag{124.13}$$

where F is an arbitrary 4-vector, S is the surface bounding the given 4-dimensional volume, and $n(x)$ is the external normal to this surface at point x . Setting

$$F_\mu(x') = K(x-x')\gamma_\mu\psi(x'),$$

we obtain

$$\begin{aligned}\frac{\partial F_\mu}{\partial x'_\mu} &= \frac{\partial K(x-x')}{\partial x'_\mu} \gamma_\mu \psi(x') + K(x-x') \gamma_\mu \frac{\partial \psi(x')}{\partial x'_\mu} = \\ &= -\frac{\partial K(x-x')}{\partial x_\mu} \gamma_\mu \psi(x') + K(x-x') \gamma_\mu \frac{\partial \psi(x')}{\partial x'_\mu}.\end{aligned}$$

From (124.10) it follows that

$$\frac{\partial K(x-x')}{\partial x_\mu} \gamma_\mu = \gamma_\mu \frac{\partial K(x-x')}{\partial x_\mu} = \hat{\nabla}_x K(x-x').$$

Making use of relations (124.1) and (124.2) we obtain

$$\frac{\partial F_\mu(x')}{\partial x'_\mu} = -(\hat{\nabla}_x + m)K(x-x')\psi(x') = i\delta^4(x-x')\psi(x').$$

Substituting this expression into the 4-dimensional integral written above and using (124.13) we obtain

$$\psi(x) = -\int_S K(x-x')\gamma_\mu\psi(x')n_\mu(x')d\sigma(x').$$

Denoting point x in terms of point 2, and point x' in terms of point 1, we can rewrite the relation obtained in the form

$$\psi(2) = \int_{t_1} K(2,1)\gamma_4\psi(1)d^3x_1 - \int_{t'_1} K(2,1')\gamma_4\psi(1')d^3x_{1'}. \quad (124.14)$$

Here two infinite space-like planes $t = t_1$ and $t = t'_1$, where $t_1 < t_2 < t'_1$, are chosen as the surface of integration. The integration over time-like surfaces can be dropped, since they are as spatially distant from point 2 as one wishes, and the function $K(2,1)$, as can be shown, decreases exponentially to zero in space-like directions as spatial distances increase indefinitely*.

The function $K(2,1)$ contains a summation only over states with a positive energy, and the function $K(2,1')$, for $t_2 < t'_1$, only over states with a negative

* See R.P.Feynman, Phys. Rev. 76 (1949) 749.

energy. Hence the first integral in (124.14) differs from zero for the components $\psi(1)$ corresponding to particles with a positive energy, and the second integral is correspondingly not equal to zero for the components $\psi(1')$ corresponding to particles with a negative energy.

We see that the wave function of a particle at point 2 of the 4-dimensional space is defined by the Green's function and by the values of $\psi(1)$ and $\psi(1')$. Analogously, setting $F_\mu = \bar{\psi}(x')\gamma_\mu K(x'-x)$ it is easy to find the expressions for the function $\bar{\psi}(2)$:

$$\bar{\psi}(2) = \int_{t_1 > t_2} \bar{\psi}(1')\gamma_4 K(1', 2) d^3x_{1'} - \int_{t_1 < t_2} \bar{\psi}(1)\gamma_4 K(1, 2) d^3x_1, \quad (124.15)$$

where the function $\bar{\psi}(x) = \psi^\dagger(x)\gamma_4$ and satisfies the equation

$$\frac{\partial \bar{\psi}}{\partial x_\mu} \gamma_\mu - m \bar{\psi} = 0$$

or

$$\bar{\psi}(\hat{\nabla} - m) = 0.$$

In this form of notation the operator $\hat{\nabla}$ acts on functions standing on its left.

The components of the wave function which correspond to negative energies E are interpreted in Feynman's theory as the amplitudes of probability of finding the particle in the positron state, i.e. in a state with positive energy $+E$ and charge $+e$. Thus the function $\psi(2)$ is given if the amplitude of the electron state $\psi(1)$, at instant of time $t_1 < t_2$, and the amplitude of the positron state $\psi(1')$ at instant of time $t_1' > t_2$, are known.

The phase factor involved in the function $K(2, 1)$ for $t_2 < t_1$ depends on time according to the law

$$\exp [iE_p(t_2 - t_1)] = \exp [-iE_p|t_2 - t_1|]$$

(see (124.12)). In other words, the time factor of the function $K(2, 1')$ depends on $(t_2 - t_1')$ in the same way as the phase factor of the wave function to particles with a positive energy. In accordance with this, positron states can be considered as states of a particle which has a positive energy but moves in the opposite direction along the time axis. To this there corresponds the fact that the state of a positron must be given at an instant of time $t_1' > t_2$ (the second integral in formula (124.14)).

We now assume that there is an external electromagnetic field. In our notation the Dirac equation in this case is written in the form

$$(\hat{\nabla} - ie\hat{A} + m)\psi = 0, \quad (124.16)$$

where $\hat{A} = A_\mu \gamma_\mu$, and $A_4 = i\varphi$ (φ is the scalar potential).

Green's function is as usual defined by the equation

$$(\hat{\nabla}_x - ie\hat{A} + m)K^A(x-x') = -i\delta(x-x').$$

The function $K^A(2,1)$, as well as the function $K(2,1)$, contains in its expansion only components corresponding to positive energies for $t_2 > t_1$ and to negative energies for $t_2 < t_1$. Relations (124.14) and (124.15) remain valid, provided that $K(2,1)$ in them is replaced by $K^A(2,1)$.

Just as in the non-relativistic case, an integral equation satisfied by the function $K^A(2,1)$ can be formulated. Namely

$$K^A(2,1) = K(2,1) - e \int K(2,3) \hat{A}(3) K^A(3,1) d^4x_3. \quad (124.17)$$

The derivation of integral equation (124.17) does not differ from that of integral equation (29.17). As we have shown in §58, an equation of such a type is conveniently solved by a method of successive approximations (see (58.5)).

§125. Green's function for a system of two particles

The expression found above for the wave function of one particle must be generalized to the case of a system of interacting particles. The simplest example of such a system is one consisting of two particles interconnected by an interaction of electromagnetic character. We note, first of all, that the Green's function of a system of two non-interacting particles is equal to the product of the Green's functions of each of the particles:

$$K(3,4;1,2) = K_a(3,1)K_b(4,2). \quad (125.1)$$

Here $K_a(3,1)$ is the Green's function of the free particle a moving from point 1 to point 3. The quantity $K_b(4,2)$ for the particle b has an analogous meaning.

In the case of two interacting particles the Green's function given by formula (125.1) can be considered as the zero order approximation with respect to the interaction. Let us now find the Green's function $K^{(1)}(3,4;1,2)$ in the first approximation with respect to the interaction. That is, we consider two charged particles which are described by the Dirac equation. In order to write the interaction operator it is convenient, following Feynman, to consider at first the non-relativistic approximation and then to carry out

the corresponding generalization to the case of relativistic particles. In the non-relativistic approximation the interaction between particles is described by the Coulomb law, and the function $K^{(1)}(3,4;1,2)$ by analogy with formula (58.3) is defined by the relation

$$K^{(1)}(3,4;1,2) = -ie^2 \int K_a(3,5) K_b(4,6) \frac{1}{r_{56}} \delta(t_{56}) K_a(5,1) K_b(6,2) dx_5^4 dx_6^4, \quad (125.2)$$

where $r_{56} = |\mathbf{r}_5 - \mathbf{r}_6|$. The meaning of the expression for $K^{(1)}(3,4;1,2)$ is easily understood if it is compared with a diagram (fig. V.31) which is interpreted as follows. Particle a moves from point 1 to point 3, passing through the intermediate point 5. The line 2-6-4 describes the motion of particle b . To the line 1-5 of the diagram there corresponds the function of motion $K_a(5,1)$, and to the line 2-6 the function $K_b(6,2)$. The interaction between the particles takes place at points 5 and 6. The dotted line corresponds to the expression $(e^2/r_{56})\delta(t_{56})$, where $r_{56} = |\mathbf{r}_5 - \mathbf{r}_6|$ is the spatial distance between points 5 and 6, and $t_{56} = t_5 - t_6$, where t_5 and t_6 are the instants of time at which particles a and b arrive at points 5 and 6. The δ -function of the time argument means that in the non-relativistic approximation one has to disregard the time lag and to consider particles at points 5 and 6 at one and the same instant of time $t = t_5 = t_6$. The lines 5-3 and 6-4 correspond to the motion of free particles after the interaction (the functions $K_a(3,5)$ and $K_b(4,6)$ in (125.2)).

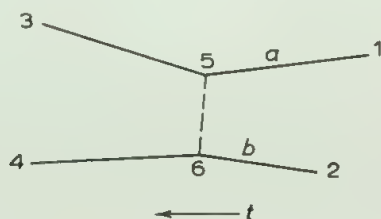


Fig. V.31

The generalization of expression (125.2) to the relativistic case involves first of all taking into account the interaction lag. At first sight it may seem that $\delta(t_{56})$ should be replaced by $\delta(t_{56} - r_{56})$, where r_{56} defines the time lag (in our notation the velocity of propagation of interaction $c = 1$). However, such a replacement would be incorrect. Indeed, the electromagnetic interaction represents the exchange of photons which have a positive energy.

However, the expansion of the δ -function in a Fourier integral contains positive as well as negative frequencies. Hence for the transition from the Coulomb interaction to the relativistic generalization taking into account the time lag, the δ -function should be replaced by the function δ_+ defined by the relation

$$\delta_+(x) = \lim_{\epsilon \rightarrow 0} \int_0^{\infty} e^{-i\omega(x-i\epsilon)} \frac{d\omega}{\pi} \equiv \lim_{\epsilon \rightarrow 0} \frac{1}{x-i\epsilon} \frac{1}{\pi i}. \quad (125.3)$$

The analogue of the δ -function defined in such a way contains the expansion only in terms of positive frequencies. Since t_{56} takes on positive as well as negative values, one takes the symmetrized combination

$$\frac{1}{2r_{56}} [\delta_+(t_{56}-r_{56}) + \delta_+(-t_{56}-r_{56})] = \delta_+(t_{56}^2-r_{56}^2) = \delta_+(-x_{56}^2).$$

This equality is immediately found by means of formula (125.3).

Furthermore, it is necessary to take into account that if the particles are moving, then in addition to the Coulomb interaction there is an electromagnetic interaction (see (25.27) of Part II). This leads to the fact that the interaction is defined by the expression $(1-\mathbf{v}_5 \cdot \mathbf{v}_6) e^2 \delta_+(-x_{56}^2)$.

We shall obtain the interaction operator if we replace the velocity vectors \mathbf{v}_5 and \mathbf{v}_6 by the operators $\boldsymbol{\alpha}_a$ and $\boldsymbol{\alpha}_b$; each of the operators acts respectively on the variables of particle a and particle b . (Indeed, the velocity operator can easily be found by the formula $\mathbf{v} = [\hat{H}, \hat{\mathbf{r}}]$. But \hat{H} (see (113.7)) is equal to $\hat{H} = i^{-1} \boldsymbol{\alpha} \nabla + \beta m$, and, commuting, we obtain $[\hat{H}, \mathbf{r}] = \boldsymbol{\alpha}$.) Then, by virtue of (113.14), one can write for the interaction operator in the relativistic case

$$(1 - \boldsymbol{\alpha}_a \cdot \boldsymbol{\alpha}_b) e^2 \delta_+(-x_{56}^2) = e^2 \beta_a \beta_b \gamma_{a\mu} \gamma_{b\mu} \delta_+(-x_{56}^2).$$

To obtain the final expression for the function $K^{(1)}(3,4;1,2)$ we have to establish the connection between relativistic and non-relativistic Green's functions. For this we compare formulae (29.3) and (124.14). We see that the following correspondence holds:

$$K_{\text{non-rel}} \rightarrow K_{\text{rel}} \beta.$$

Thus for particles by the Dirac equation we have

$$K^{(1)}(3,4;1,2) \beta_a \beta_b = -ie^2 \int K_a(3,5) K_b(4,6) \gamma_{a\mu} \gamma_{b\mu} \delta_+(-x_{56}^2) \times \\ \times K_a(5,1) K_b(6,2) \beta_a \beta_b d^4 x_5 d^4 x_6. \quad (125.4)$$

Multiplying (125.4) from the left by $\beta_a \beta_b$, we find finally

$$\begin{aligned}
 K^{(1)}(3,4;1,2) &= -ie^2 \int K_a(3,5)K_b(4,6)\gamma_{a\mu}\gamma_{b\mu}\delta_+(-x_{56}^2) \times \\
 &\times K_a(5,1)K_b(6,2)d^4x_5d^4x_6 = \\
 &= e^2 \int K_a(3,5)K_b(4,6)\gamma_{a\mu}D(-x_{56}^2)\gamma_{b\mu}K_a(5,1)K_b(6,2)d^4x_5d^4x_6.
 \end{aligned}
 \tag{125.5}$$

The function $D(-x_{56}^2) = -i\delta_+(-x_{56}^2)$ is usually called the propagation function of the virtual photon. Thus we see that, taking into account relativistic effects, the diagram in fig. V.31 can be interpreted as follows. The functions K correspond to solid (electron) lines, the function D corresponds to the dotted line, and to the vertices there correspond the matrices $e\gamma_{a\mu}$ and $e\gamma_{b\mu}$.

All calculations in Feynman's theory are substantially simplified if they are carried out in the momentum representation. The form of the function K in the p -representation is given by formula (124.8). There remains to be determined the Fourier component of the function δ_+ . We shall show that the following relation holds:

$$\delta_+(-x^2) = \frac{1}{4\pi^3} \int \frac{e^{ikx}}{k^2 - i\epsilon} d^4k, \tag{125.6}$$

where ϵ is an infinitesimal quantity. It defines the rule for circumventing poles. We can convince ourselves of the validity of relation (125.6) by calculating the integral on the right directly:

$$\int \frac{e^{ikx}}{k^2 - i\epsilon} d^4k = \int e^{ik \cdot r} d^3k \int_{-\infty}^{\infty} \frac{e^{-ik_0x_0}}{k^2 - k_0^2 - i\epsilon} dk_0.$$

We assume, for example, that $x_0 > 0$. Closing the contour of integration below and finding the residues, we obtain

$$2\pi i \int \frac{e^{ik \cdot r}}{2|k|} e^{-ik|x_0|} d^3k.$$

Writing d^3k in the form $k^2 dk d\Omega$, integrating over angles and taking into account (125.3) we obtain the relation sought.

§126. Feynman diagrams

We shall now consider the rules for calculation of the probabilities of transition from one state into another by means of the mathematical apparatus presented in the preceding sections. For simplicity we shall first con-

sider one particle (for example, an electron) which makes a transition from one state into another under the action of an external electromagnetic field. Let the electron at the initial instant of time $t = t_1$ be in the state $\psi(\mathbf{r}_1, t_1) = \psi(1)$, and at the instant of time $t = t_2$ let it be in the state $\psi(\mathbf{r}_2, t_2) = \psi(2)$ corresponding to a positive energy. The probability of transition into a particular state $\psi_n(\mathbf{r}_2, t_2)$ is, as always, defined by the square of the modulus of the corresponding amplitude of the expansion of the function $\psi(\mathbf{r}_2, t_2)$ in terms of the function $\psi_n(\mathbf{r}_2, t_2)$

$$M = \int \psi_n^\dagger(\mathbf{r}_2, t_2) \psi(\mathbf{r}_2, t_2) d^3x_2. \quad (126.1)$$

Expressing the function $\psi(\mathbf{r}_2, t_2)$ in terms of Green's function according to formula (124.14), we obtain

$$M = \int \psi_n^\dagger(\mathbf{r}_2, t_2) K^A(\mathbf{r}_2, t_2; \mathbf{r}_1, t_1) \beta \psi(\mathbf{r}_1, t_1) d^3x_1 d^3x_2.$$

In place of the Green's function K^A we can write its expansion in a series of successive approximations. We then obtain an expression for the transition amplitude M in the form of a perturbation theory series. Thus, for example, the transition amplitude in the first approximation of perturbation theory is equal to

$$M^{(1)} = -e \int \psi_n^\dagger(2) K(2, 3) \hat{A}(3) K(3, 1) \beta \psi(1) d^3x_1 d^3x_2 d^4x_3. \quad (126.2)$$

This expression can be written in a more compact form, using the relations

$$\begin{aligned} \psi(3) &= \int K(3, 1) \beta \psi(1) d^3x_1, \\ \bar{\psi}_n(3) &= \int \bar{\psi}_n(2) \beta K(2, 3) d^3x_2. \end{aligned}$$

Then for the transition amplitude we have

$$M^{(1)} = -e \int \bar{\psi}_n(3) \hat{A}(3) \psi(3) d^4x_3. \quad (126.3)$$

It is easy to obtain, in an analogous way, the second approximation of the transition amplitude

$$M^{(2)} = (-e)^2 \int \bar{\psi}_n(3) \hat{A}(3) K(3, 4) \hat{A}(4) \psi(4) d^4x_3 d^4x_4. \quad (126.4)$$

If the initial and final states are described by plane waves, then formulae (126.3) and (126.4) are conveniently rewritten in the momentum representation. Setting

$$\begin{aligned} \bar{\psi}_n(3) &= \bar{u}(\mathbf{p}_2) e^{-ip_2 x_3}, \\ \psi(3) &= u(\mathbf{p}_1) e^{ip_1 x_3} \end{aligned}$$

and using the Fourier representation of the operator \hat{A}

$$\hat{A}(3) = \int \hat{a}(k) e^{ikx_3} d^4k, \quad (126.5)$$

we obtain for the transition amplitude of the first order (126.3)

$$\begin{aligned} M^{(1)} &= -e \int d^4x_3 \int e^{-ip_2x_3+ikx_3+ip_1x_3} \bar{u}(p_2) \hat{a}(k) u(p_1) d^4k = \\ &= -e(2\pi)^4 \int \bar{u}(p_2) \hat{a}(k) \delta^4(k+p_1-p_2) u(p_1) d^4k = \\ &= -e(2\pi)^4 \bar{u}(p_2) \hat{a}(p_2-p_1) u(p_1). \end{aligned} \quad (126.6)$$

For the transition amplitude of the second order we have, correspondingly,

$$\begin{aligned} M^{(2)} &= (-e)^2 \int \bar{u}(p_2) \hat{a}(k_1) S(p) \hat{a}(k) u(p_1) d^4p d^4k d^4k_1 \times \\ &\times \int e^{-ip_2x_3+ik_1x_3+ip(x_3-x_4)+ikx_4+ip_1x_4} d^4x_3 d^4x_4 = \\ &= e^2(2\pi)^8 \int \bar{u}(p_2) \hat{a}(k_1) \delta^4(k_1+p-p_2) S(p) \hat{a}(k) \times \\ &\times \delta^4(p_1+k-p) u(p_1) d^4p d^4k d^4k_1 = \\ &= e^2(2\pi)^8 \int \bar{u}(p_2) \hat{a}(p_2-p_1-k) \times \\ &\times \frac{1}{i(2\pi)^4 [i(\hat{p}_1 + \hat{k}) + m]} \hat{a}(k) u(p_1) d^4k. \end{aligned} \quad (126.7)$$

Formulae (126.6) and (126.7) can be associated with pictorial representations, called Feynman diagrams. As will be shown below, to each line and to each crossing of lines (called a vertex) in the Feynman diagram there corresponds a definite factor in the transition amplitude. In the case of complex processes such diagrams make it possible to simplify the construction of expressions for the transition amplitudes. In a Feynman diagram we represent the states of electrons and positrons by solid lines, and the states of the electromagnetic field by dotted lines. The arrows on the lines show the order of writing the terms of the transition amplitude. To an increase in time there corresponds the motion of the particle from the right to the left.

Let us consider the simplest Feynman diagram (fig. V.32) corresponding to the following process: an electron with momentum p_1 was scattered by an external electromagnetic field and made a transition into a new state with momentum p_2 . The probability amplitude of this transition is given by formula (126.6). In fig. V.32 the free electron with momentum p_1 is represented by the solid line AB. This straight line corresponds to the first factor in the transition amplitude $M^{(1)}$ (the factors are numbered from the right to the left), the bispinor $u(p_1)$. At point B the electron is scattered by the electromagnetic field represented by the dotted line. The crossing of the solid and dotted lines (the vertex) in the Feynman diagram corresponds to the operator

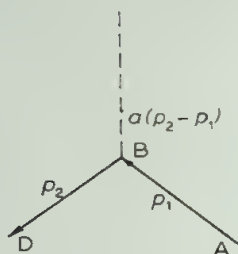


Fig. V.32

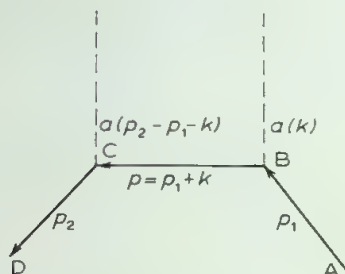


Fig. V.33

$-e\hat{a}(k)$ in the transition amplitude $M^{(1)}$ multiplied by the δ -function of the momenta of all three particles. The electron with momentum \mathbf{p}_2 is represented by the straight line BD. In the amplitude $M^{(1)}$ the bispinor $\bar{u}(\mathbf{p}_2)$ corresponds to it.

We see that the order of the process with respect to the charge e is defined by the number of vertices in the Feynman diagram. This is particularly clearly seen from the consideration of the Feynman diagram for a process of the second order (fig. V.33). This diagram corresponds to the process of electron scattering in the second approximation of perturbation theory. The line AB (called the external line) represents the motion of the free electron. To it there corresponds the bispinor $u(\mathbf{p}_1)$ in the transition amplitude $M^{(2)}$. The scattering of the electron takes place at vertex B. In the transition amplitude $M^{(2)}$ there corresponds to vertex B the factor $-e\hat{a}(k)$ and the momentum δ -function $\delta^4(p_1 + k - p)$. The line BC joining the two vertices is called the internal line. To it there corresponds in $M^{(2)}$ the factor S , the Fourier component of the Green's function defined by formula (124.8). The external field acts at vertex C. To the vertex C there corresponds in $M^{(2)}$ the operator $-e\hat{a}(k_1)$ and the δ -function $\delta^4(k_1 + p - p_2)$. The external line CD represents the motion of the electron with momentum \mathbf{p}_2 . To the line CD in $M^{(2)}$ there corresponds the bispinor $\bar{u}(\mathbf{p}_2)$. Since the law of conservation of 4-momentum in the transition from state \mathbf{p}_1 into state \mathbf{p}_2 is fulfilled for arbitrary values of the wave vector k , the integration is carried out over the vector k (or k_1). The value of the numerical factor in the expression for the amplitude is determined by the number of δ -functions involved in it, which is equal to the number of vertices. Each vertex brings into $M^{(2)}$ a factor of $(2\pi)^4$.

The Feynman diagram for processes involving positrons can be constructed in exactly the same way. For example, the diagram in fig. V.33 also describes

positron scattering in the second approximation of perturbation theory. Since in Feynman's theory the positron is considered as an electron moving backward in time, this diagram defines the transition amplitude of the positron from a state with momentum $-p_2$ into a state with momentum $-p_1$. For the quantity $M^{(2)}$ in this case we have

$$M^{(2)} = \frac{1}{i} e^2 (2\pi)^4 \int \bar{v}(p_2) \hat{a}(p_2 - p_1 - k) \frac{1}{i(\hat{p}_1 + \hat{k}) + m} \hat{a}(k) v(p_1) d^4 k. \quad (126.8)$$

Here v is the Dirac bispinor corresponding to a state with a negative energy.

The relations derived make it possible to consider processes associated with the emission and absorption of free electrons in addition to scattering processes in an external electromagnetic field. For this the operator \hat{A} in the general formulae (126.3) and (126.4) must correspond to the pole of one emitted or one absorbed photon. In accordance with formulae (102.3) and (102.5), there corresponds to the pole of the emitted photon the vector potential

$$A_\mu = (2\pi/\omega)^{\frac{1}{2}} e_\mu e^{-ikx}, \quad (126.9)$$

and to the absorbed photon

$$A_\mu = (2\pi/\omega)^{\frac{1}{2}} e_\mu e^{ikx}. \quad (126.10)$$

Here e_μ is the polarization vector, and k denotes a four-dimensional wave vector. Since only one photon is absorbed or emitted, the matrix elements of the operators \hat{a} and \hat{a}^\dagger are equal to one. The diagrams shown in fig. V.34 describe processes involving two free photons. Thus, for example, these diagrams describe the process of Compton scattering, i.e. the scattering of a photon by a free electron. In this process the photon before scattering had wave vector k_1 , and after scattering k_2 . In the diagram shown in fig. V.34a

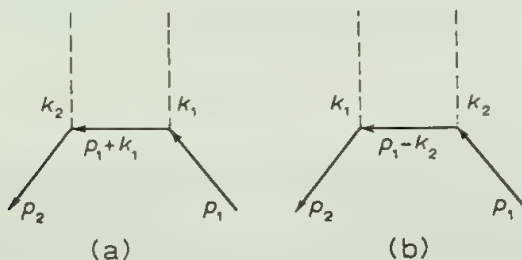


Fig. V.34

this corresponds to the fact that the photon is first absorbed and then emitted by the electron. The diagram shown in fig. V.34b also corresponds to the same process, where first the photon \mathbf{k}_2 is emitted and then the photon \mathbf{k}_1 is absorbed. Of course, the words 'first' and 'then' refer only to the order of writing the factors in the transition amplitude and have no other, physical meaning.

The Feynman diagrams shown in fig. V.34 make it possible to write immediately the transition amplitude without having to carry out each time special calculations of the type carried out in deriving formulae (126.6) and (126.7). The total transition amplitude is defined by the sum of the amplitudes corresponding to diagrams V.34a and V.34b.

It has the form

$$M = \frac{e^2}{i} (2\pi)^4 \left[\bar{u}(\mathbf{p}_2) \left(\frac{2\pi}{\omega_2} \right)^{\frac{1}{2}} \hat{e}_2 \frac{1}{i(\hat{p}_1 + \hat{k}_1) + m} \left(\frac{2\pi}{\omega_1} \right)^{\frac{1}{2}} \hat{e}_1 u(\mathbf{p}_1) + \right. \\ \left. + \bar{u}(\mathbf{p}_2) \left(\frac{2\pi}{\omega_1} \right)^{\frac{1}{2}} \hat{e}_1 \frac{1}{i(\hat{p}_1 - \hat{k}_2) + m} \left(\frac{2\pi}{\omega_2} \right)^{\frac{1}{2}} \hat{e}_2 u(\mathbf{p}_1) \right] \delta^4(p_1 + k_1 - p_2 - k_2), \quad (126.11)$$

where $\hat{e}_1 = e_{1\mu} \gamma_\mu$, $\hat{e}_2 = e_{2\mu} \gamma_\mu$, and e_1 and e_2 are the polarization vectors of the photon before and after scattering.

The appearance of the δ -function is easily understood if it is taken into account that in the given case expressions of the type (126.9) and (126.10), containing no integration over k , are substituted for the external field operator of the form of (126.5) in the expression of the type (126.4). Thus in an expression of the type (126.7) there will be no integration over k and k_1 . After the integration over p there will remain one δ -function, expressing the law of conservation of energy and momentum in the Compton process.

Diagrams of such a type describe, for example, the process of annihilation of an electron with momentum \mathbf{p}_1 and of a positron with momentum $-\mathbf{p}_2$. Two photons with momenta \mathbf{k}_1 and \mathbf{k}_2 are produced in the annihilation (fig. V.35). The amplitude of the two-photon annihilation of the pair, according to the same rules, is written in the form

$$M = \frac{1}{i} e^2 (2\pi)^5 \frac{1}{(\omega_1 \omega_2)^{\frac{1}{2}}} \times \\ \times \left[\bar{v}(\mathbf{p}_2) \left(\hat{e}_1 \frac{1}{i(\hat{p}_1 - \hat{k}_2) + m} \hat{e}_2 + \hat{e}_2 \frac{1}{i(\hat{p}_1 - \hat{k}_1) + m} \hat{e}_1 \right) u(\mathbf{p}_1) \right] \times \\ \times \delta^4(p_1 - p_2 - k_1 - k_2). \quad (126.12)$$

As another example let us consider the electron bremsstrahlung, i.e. the

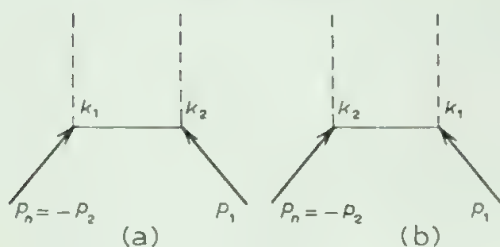


Fig. V.35

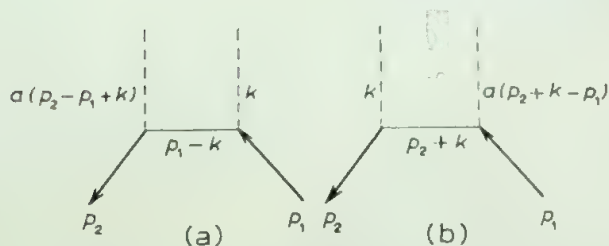


Fig. V.36

radiation arising when a fast electron traverses the field of a nucleus. The diagram corresponding to this process is shown in fig. V.36. An electron with momentum \mathbf{p}_1 scattered by an external field a emits a photon of momentum \mathbf{k} and polarization \mathbf{e} and makes a transition into a state with momentum \mathbf{p}_2 . In this case two processes, shown in the diagrams of fig. V.36a and b, are possible. The total transition amplitude, in accordance with the rules presented, is of the form

$$M = \frac{e^2}{i} \frac{(2\pi)^{\frac{3}{2}}}{\omega^{\frac{1}{2}}} \bar{u}(\mathbf{p}_2) \left[\hat{a}(\mathbf{p}_2 - \mathbf{p}_1 + \mathbf{k}) \frac{1}{i(\hat{\mathbf{p}}_1 - \hat{\mathbf{k}}) + m} \hat{\mathbf{e}} + \right. \\ \left. + \hat{\mathbf{e}} \frac{1}{i(\hat{\mathbf{p}}_2 + \hat{\mathbf{k}}) + m} \hat{a}(\mathbf{p}_2 + \mathbf{k} - \mathbf{p}_1) \right] u(\mathbf{p}_1). \quad (126.13)$$

Let us now consider processes associated with the interaction of two particles, for example the scattering of an electron by a μ -meson. The transition amplitude is again calculated by expanding the wave function of the system in terms of the products of the wave functions of the free particles. The wave function of a system of two particles is defined by the Green's

function (125.5). These calculations lead to graphs which are plotted according to the same principles as for one particle. In the first approximation of perturbation theory there corresponds to the scattering process the Feynman diagram shown in fig. V.37.

Here the solid lines AB and CD correspond to the motion of free particles with momenta p_1 and p_2 . The electromagnetic interaction between particles amounts to photon exchange. A virtual photon, which is emitted by the second particle (vertex D), is absorbed at vertex B. Lines BE and DF correspond to the motion of the particles after the interaction with momenta p_3 and p_4 . The transition amplitude is constructed according to the usual rules and has the form

$$M = e^2 (2\pi)^8 \bar{u}_a(p_3) \bar{u}_b(p_4) \gamma_{a\mu} D_f(p_1 - p_3) \times \\ \times \gamma_{b\mu} u_a(p_1) u_b(p_2) \delta^4(p_3 + p_4 - p_1 - p_2). \quad (126.14)$$

Here D_f is the Fourier component of the propagation function of the virtual photon D, which, according to (125.5) and (125.6), is given by the formula

$$D_f(k) = -\frac{i}{4\pi^3} \frac{1}{k^2}. \quad (126.15)$$

Let us dwell briefly on corrections which arise in higher approximations of perturbation theory. The diagrams corresponding to these corrections must,

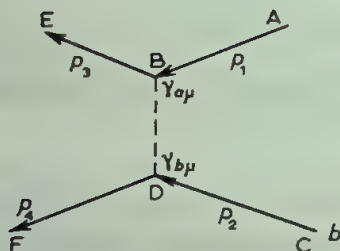


Fig. V.37.

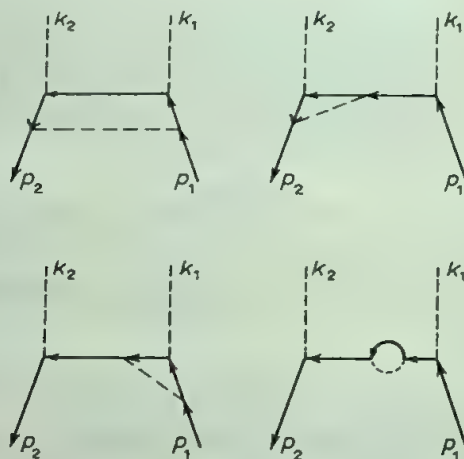


Fig. V.38

naturally, contain a larger number of vertices in comparison with the corresponding basic diagram (to each vertex there corresponds the smallness parameter e). Namely, it is the number of vertices that determines the order of smallness of the correction considered.

Let us consider, for example, diagrams corresponding to the next approximation in the theory of the Compton effect. As is easily understood, to the diagram in fig. V.34a there correspond the corrections in fig. V.38, and analogously for the diagram in fig. V.34b. All these diagrams differ from the initial diagram (fig. V.34) by the presence of an internal photon line. This line corresponds, as we have already seen, to the emission and absorption of a virtual quantum. Hence these corrections are usually called radiative corrections. The calculation of these corrections involves particular difficulties and requires the so-called renormalization method. We shall not dwell on these problems*.

§ 127. The Compton effect

To illustrate the technique of calculation of cross sections in Feynman's theory we shall consider the theory of the Compton effect in more detail. The transition amplitude for this process has already been obtained by means of a Feynman diagram (fig. V.34) in the form of (126.11). Separating the δ -function, we write the transition amplitude in the form

$$M = M_{21} \delta^4(p_1 + k_1 - p_2 - k_2), \quad (127.1)$$

where

$$M_{21} = \frac{ie^2(2\pi)^5}{(\omega_1 \omega_2)^{\frac{1}{2}}} \bar{u}(p_2) \left[\hat{e}_2 \frac{i(\hat{p}_1 + \hat{k}_1) - m}{(\hat{p}_1 + \hat{k}_1)^2 + m^2} \hat{e}_1 + \hat{e}_1 \frac{i(\hat{p}_1 - \hat{k}_2) - m}{(\hat{p}_1 - \hat{k}_2)^2 + m^2} \hat{e}_2 \right] u(p_1).$$

The probability of the Compton effect is given by the formula

$$P'_{21} = |M_{21}|^2 (\delta^4(p_1 + k_1 - p_2 - k_2))^2. \quad (127.2)$$

In order to eliminate the square of the δ -function it is convenient to make use of the definition of the 4-dimensional δ -function

$$\delta^4(p) = \frac{1}{(2\pi)^4} \int e^{ipx} d^4x.$$

* See, for example, A.I.Akhiezer and V.B.Berestetskii, *Quantum electrodynamics* (Interscience Publ., New York, 1965).

At the point $p = 0$, which, as is seen from (127.2), is the only one playing an important role, the 4-dimensional integral is equal to $VT/(2\pi)^4$, where V is the normalization volume, and T is the duration of the process. Choosing V to be unity, we have for the transition probability P_{21} per unit time

$$P_{21} = \frac{1}{T} P'_{21} = \frac{1}{(2\pi)^4} |M_{21}|^2 \delta^4(p_1 + k_1 - p_2 - k_2). \quad (127.3)$$

The final state of the system is defined by the momentum of the electron \mathbf{p}_2 and of the scattered photon \mathbf{k}_2 . The number of final states in the interval of momenta $d\mathbf{p}_2$ and $d\mathbf{k}_2$ is given by the usual relation $d\mathbf{p}_2 d\mathbf{k}_2 / (2\pi)^6$. The probability of transition into the interval of final states $d\mathbf{p}_2 d\mathbf{k}_2$ is written in the form

$$dW_{21} = \frac{1}{(2\pi)^4} |M_{21}|^2 \delta^4(p_1 + k_1 - p_2 - k_2) \frac{d\mathbf{p}_2 d\mathbf{k}_2}{(2\pi)^6}. \quad (127.4)$$

The 4-dimensional δ -function expresses the law of conservation of energy and momentum

$$p_1 + k_1 = p_2 + k_2. \quad (127.5)$$

From this relation it is easy to define the frequency of the scattered photon as a function of the scattering angle ϑ , i.e. the angle between the vectors \mathbf{k}_1 and \mathbf{k}_2

$$p_1^2 + k_1^2 + 2p_1 k_1 = p_2^2 + k_2^2 + 2p_2 k_2,$$

but

$$p_1^2 = p_1^2 - E_1^2 = -m^2 = p_2^2,$$

$$k_1^2 = k_1^2 - \omega_1^2 = k_2^2 = 0$$

and, consequently, $p_1 k_1 = p_2 k_2$. Making use of (127.5) we have

$$p_1 k_1 = p_1 k_2 + k_1 k_2.$$

We assume for simplicity that the electron was initially at rest: $\mathbf{p}_1 = 0$, $E_1 = m$. After a simple calculation (see (17.11) of Part II) we find

$$\omega_2 = \frac{\omega_1}{1 + m^{-1} \omega_1 (1 - \cos \vartheta)}. \quad (127.6)$$

Integrating (127.4) over the three momentum components we obtain

$$dW_{21} = \frac{1}{(2\pi)^{10}} |M_{21}|^2 \delta(m + \omega_1 - E_2 - \omega_2) \omega_2^2 d\omega_2 d\Omega. \quad (127.7)$$

We integrate this expression over frequency ω_2 . It should then be recalled that the energy E_2 is also a function of ω_2 ,

$$E_2 = [m^2 + p_2^2]^{\frac{1}{2}} = [m^2 + (k_1 - k_2)^2]^{\frac{1}{2}} = [m^2 + \omega_1^2 + \omega_2^2 - 2\omega_1 \omega_2 \cos \vartheta]^{\frac{1}{2}}.$$

We introduce the new variable $y = E_2 + \omega_2$

$$dW_{21} = \frac{1}{(2\pi)^{10}} |M_{21}|^2 \delta(m + \omega_1 - y) \omega_2^2 \frac{1}{\partial y / \partial \omega_2} dy d\Omega. \quad (127.8)$$

Integrating with respect to y we have

$$dW_{21} = \frac{1}{(2\pi)^{10}} |M_{21}|^2 \omega_2^2 \left. \frac{1}{\partial y / \partial \omega_2} \right|_{\omega_2 + E_2 = m + \omega_1} d\Omega, \quad (127.9)$$

but

$$\frac{\partial y}{\partial \omega_2} = \frac{\partial(E_2 + \omega_2)}{\partial \omega_2} = 1 + \frac{1}{E_2} (\omega_2 - \omega_1 \cos \vartheta). \quad (127.10)$$

Since the value of ω_2 satisfying the energy conservation law is used, then substituting (127.6) into (127.10), we find for the transition probability (127.9)

$$dW_{21} = \frac{1}{(2\pi)^{10}} \frac{\omega_2^3 E_2}{m \omega_1} |M_{21}|^2 d\Omega. \quad (127.11)$$

We shall obtain the cross section for the process by dividing the transition probability by the incident photon flux density. For a normalization such that there is 1 photon in volume $V = 1$ the flux density is numerically equal to the velocity of light c . In the system of units that we have chosen $c = 1$ and the cross section turns out to be numerically equal to dW_{21} . We also set (see (127.1))

$$M_{21} = ie^2 \frac{(2\pi)^5}{2m(\omega_1 \omega_2)^{\frac{1}{2}}} (\bar{u}(p_2) \hat{Q} u(p_1)),$$

where

$$Q = \left[\hat{e}_2 \frac{i(\hat{p}_1 + \hat{k}_1) - m}{(\hat{p}_1 + \hat{k}_1)^2 + m^2} \hat{e}_1 + \hat{e}_1 \frac{i(\hat{p}_1 - \hat{k}_2) - m}{(\hat{p}_1 - \hat{k}_2)^2 + m^2} \hat{e}_2 \right] 2m.$$

The expression for the operator Q can be somewhat simplified if it is taken into account that

$$(\hat{p}_1 + \hat{k}_1)^2 + m^2 = p_1^2 + k_1^2 + 2p_1 k_1 + m^2 = 2p_1 k_1 = -2\omega_1 m.$$

Analogously for the denominator of the second fraction one can write

$$(\hat{p}_1 - \hat{k}_2)^2 + m^2 = p_1^2 + k_2^2 - 2p_1 k_2 + m^2 = -2p_1 k_2 = 2\omega_2 m.$$

Then for \hat{Q} we have

$$\hat{Q} = \hat{e}_1 \frac{i(\hat{p}_1 - \hat{k}_2) - m}{\omega_2} \hat{e}_2 - \hat{e}_2 \frac{i(\hat{p}_1 + \hat{k}_1) - m}{\omega_1} \hat{e}_1. \quad (127.12)$$

Correspondingly the differential cross section is equal to

$$d\sigma = \frac{1}{4} r_0^2 \frac{\omega_2^2 E_2}{m\omega_1^2} |\bar{u}(\mathbf{p}_2) \hat{Q} u(\mathbf{p}_1)|^2 d\Omega, \quad (127.13)$$

where $r_0 = e^2/m$ is the classical radius of the electron.

The expression obtained describes a process in which the electron and photon in the initial and final states have definite polarizations. If the electrons in the initial state are not polarized and we are not interested in the polarization in the final state, then the cross section must be averaged over the spin states of the electron in the initial state and summed over the final spin states. Consequently we have to determine the quantity

$$\begin{aligned} \overline{d\sigma} &= \frac{1}{2} \sum_{\sigma_2} \sum_{\sigma_1} \frac{1}{4} r_0^2 \frac{\omega_2^2 E_2}{m\omega_1^2} |\bar{u}_{\sigma_2}(\mathbf{p}_2) \hat{Q} u_{\sigma_1}(\mathbf{p}_1)|^2 d\Omega = \\ &= \frac{1}{8} r_0^2 \frac{\omega_2^2 E_2}{m\omega_1^2} \sum_{\sigma_1, \sigma_2} (u_{\sigma_2}^\dagger \beta \hat{Q} u_{\sigma_1}) (u_{\sigma_1}^\dagger \hat{Q}^\dagger \beta u_{\sigma_2}), \end{aligned} \quad (127.14)$$

where u_{σ_1} and u_{σ_2} are states with a definite polarizations. For example, $(\hat{\mathbf{S}} \cdot \mathbf{p}_2) u_{\sigma_2} = \sigma_2 |\mathbf{p}_2| u_{\sigma_2}$, i.e. σ_2 is the spin component along the direction of motion, $\sigma_2 = \pm \frac{1}{2}$. The operator $\hat{\mathbf{S}}$ is defined in § 117.

The sum over σ_1 involved in (127.14) is conveniently rewritten in the form

$$\sum_{\sigma_1} (u_{\sigma_2}^\dagger \beta \hat{Q} u_{\sigma_1}) (u_{\sigma_1}^\dagger \hat{Q}^\dagger \beta u_{\sigma_2}) = \sum_{\sigma_1} (\sigma_2 |\beta \hat{Q}| \sigma_1) (\sigma_1 |\hat{Q}^\dagger \beta| \sigma_2). \quad (127.15)$$

The summation is carried out over two spin states σ_1 with a positive energy. If the summation were also carried out over states with a negative energy, i.e. over all possible states (for given momentum), then expression (127.15) would be considerably simplified and would represent the matrix

element $\langle \sigma_2 | \beta \hat{Q} \hat{Q}^\dagger \beta | \sigma_2 \rangle$ (in correspondence with the rule of multiplication of matrices (45.6)).

In order to extend the summation to all four intermediate states, the following method is used in calculations of such a type: the auxiliary operator \hat{R} , called the projection operator, is introduced

$$\hat{R} = \frac{\hat{H} + |E|}{2|E|} = \frac{\alpha \cdot \hat{\mathbf{p}} + \beta m + |E|}{2|E|}. \quad (127.16)$$

The action of this operator on the functions u and v which correspond to positive and negative energies is defined by the equations

$$\frac{\hat{H} + |E|}{2|E|} u = u; \quad \frac{\hat{H} + |E|}{2|E|} v = 0. \quad (127.17)$$

Replacing the function u_{σ_1} in (127.15) by the function

$$\frac{\alpha \cdot \hat{\mathbf{p}}_1 + \beta m + |E_1|}{2|E_1|} u_{\sigma_1}$$

we can formally extend the summation also to states with a negative energy, since in the presence of the projection operator, in correspondence with (127.17), they give no contribution to the result.

In place of the matrices α_k and β we introduce the matrices γ_μ

$$\alpha_k = i\gamma_4\gamma_k \quad (k=1,2,3), \quad \beta = \gamma_4.$$

Substituting the matrices γ_μ into (127.16) we obtain

$$\hat{R} = -\frac{1}{2E} (i\gamma_u p_u - m)\gamma_4 = -\frac{1}{2E} (i\hat{\mathbf{p}} - m)\gamma_4. \quad (127.18)$$

The sum (127.15) can be rewritten in the form

$$\sum_{\sigma_1, E_1} (u_{\sigma_2}^\dagger \gamma_4 \hat{Q} \hat{R}_1 u_{\sigma_1}) (u_{\sigma_1}^\dagger \hat{Q}^\dagger \gamma_4 u_{\sigma_2}) = (u_{\sigma_2}^\dagger \gamma_4 \hat{Q} \hat{R}_1 \hat{Q}^\dagger \gamma_4 u_{\sigma_2}). \quad (127.19)$$

Further, we calculate the sum over the components of the spin in the final state of the electron σ_2 . Here it is also convenient to pass to the summation over all four states, introducing the projection operator \hat{R}_2

$$S = \sum_{\sigma_2, E_2} (u_{\sigma_2}^\dagger \gamma_4 \hat{Q} \hat{R}_1 \hat{Q}^\dagger \gamma_4 \hat{R}_2 u_{\sigma_2}). \quad (127.20)$$

The summation is carried out over states with a positive energy as well as

over states with a negative energy. We see that expression (127.20) represents the sum of diagonal matrix elements, i.e.

$$S = \text{Tr} (\gamma_4 \hat{Q} \hat{R}_1 \hat{Q}^\dagger \gamma_4 \hat{R}_2) = \text{Tr} (\hat{Q} \hat{R}_1 \hat{Q}^\dagger \gamma_4 \hat{R}_2 \gamma_4), \quad (127.21)$$

since a cyclic permutation of matrices can be carried out under the sign Tr , as can easily be checked directly. Substituting this expression into (127.14) and taking into account that $\gamma_4^2 = 1$, $\hat{\mathbf{p}}_1 = 0$, $E_1 = m$, we obtain

$$d\sigma = \frac{1}{32} r_0^2 \frac{\omega_2^2}{m^2 \omega_1^2} \text{Tr} [\hat{Q} (i\hat{\mathbf{p}}_1 - m) \gamma_4 \hat{Q}^\dagger \gamma_4 (i\hat{\mathbf{p}}_2 - m)]. \quad (127.22)$$

For any operator of the form of \hat{A} , the fourth component of which is imaginary while three components are real, the following equality holds:

$$\gamma_4 \hat{A}^\dagger \gamma_4 = -\hat{A}. \quad (127.23)$$

For the product of operators we have correspondingly

$$\gamma_4 \hat{A}^\dagger \hat{B}^\dagger C^\dagger \gamma_4 = \gamma_4 \hat{A}^\dagger \gamma_4 \gamma_4 \hat{B}^\dagger \gamma_4 \gamma_4 C^\dagger \gamma_4 = (-\hat{A})(-\hat{B})(-\hat{C}). \quad (127.24)$$

Then the expression $\gamma_4 \hat{Q}^\dagger \gamma_4$ is rewritten in the form

$$\begin{aligned} \gamma_4 \{ \omega_1^{-1} \hat{e}_1^\dagger [i(\hat{\mathbf{p}}_1 + \hat{\mathbf{k}}_1)^\dagger + m] \hat{e}_2^\dagger - \omega_2^{-1} \hat{e}_2^\dagger [i(\hat{\mathbf{p}}_1 - \hat{\mathbf{k}}_2)^\dagger + m] \hat{e}_1^\dagger \} \gamma_4 = \\ = \{ \omega_1^{-1} \hat{e}_1 [-i(\hat{\mathbf{p}}_1 + \hat{\mathbf{k}}_1) + m] \hat{e}_2 + \omega_2^{-1} \hat{e}_2 [i(\hat{\mathbf{p}}_1 - \hat{\mathbf{k}}_2) - m] \hat{e}_1 \}. \end{aligned} \quad (127.25)$$

We now sum the cross section over the final states of the photon and average over the initial states. In order to calculate the cross section for the case where the incident photon is polarized along axis 1 and the scattered photon along axis 2, one has to substitute the values $\hat{e}_1 = \gamma_1$ and $\hat{e}_2 = \gamma_2$ in the expression (127.12) for \hat{Q} . Since, however, the incident photons are not polarized and we are not interested in the polarization of the scattered photons, we have to substitute γ_ν for \hat{e}_1 and γ_μ for \hat{e}_2 and to sum over all values of the indices ν and μ . We then have

$$\begin{aligned} d\sigma = \frac{r_0^2 \omega_2^2}{32 m^2 \omega_1^2} \times \\ \times \text{Tr} \{ [\omega_2^{-1} \gamma_\nu (i(\hat{\mathbf{p}}_1 - \hat{\mathbf{k}}_2) - m) \gamma_\mu - \omega_1^{-1} \gamma_\mu (i(\hat{\mathbf{p}}_1 + \hat{\mathbf{k}}_1) - m) \gamma_\nu] (i\hat{\mathbf{p}}_1 - m) \times \\ \times [\omega_2^{-1} \gamma_\mu (i(\hat{\mathbf{p}}_1 - \hat{\mathbf{k}}_2) - m) \gamma_\nu - \omega_1^{-1} \gamma_\nu (i(\hat{\mathbf{p}}_1 + \hat{\mathbf{k}}_1) - m) \gamma_\mu] (i\hat{\mathbf{p}}_2 - m) \}, \end{aligned} \quad (127.26)$$

The summation is carried out over the twice repeated indices μ and ν .

Although a free photon can be polarized in two directions perpendicular to the direction of motion, the summation over μ and ν can actually be carried out over all four values. This is associated with the fact that there are

no real photons polarized in the direction of motion and along the time axis, and taking them into account formally does not change the final result*.

Relation (127.26) is conveniently rewritten in the form

$$d\sigma = \frac{r_0^2 \omega_2^2}{32m^2 \omega_1^2} (\text{Tr } F_1 + \text{Tr } F_2), \quad (127.27)$$

where

$$F_1 = [\omega_2^{-1} \gamma_\nu (i\hat{q}_2 - m) \gamma_\mu - \omega_1^{-1} \gamma_\mu (i\hat{q}_1 - m) \gamma_\nu] (i\hat{p}_1 - m) \times \\ \times [\gamma_\mu \omega_2^{-1} (i\hat{q}_2 - m) \gamma_\nu] (i\hat{p}_2 - m), \quad (127.28)$$

$$q_1 = p_1 + k_1, \quad q_2 = p_1 - k_2.$$

The expression for F_2 is obtained from F_1 by the replacement

$$q_1 \rightarrow q_2, \quad q_2 \rightarrow q_1; \quad \omega_1 \rightarrow -\omega_2, \quad \omega_2 \rightarrow -\omega_1.$$

For further calculations it is convenient to use Feynman's formulae, which are easily checked by direct calculation:

$$\begin{aligned} \gamma_\nu \gamma_\nu &= 4, \\ \gamma_\nu \hat{A} \gamma_\nu &= -2\hat{A}, \\ \gamma_\nu \hat{A}_1 \hat{A}_2 \gamma_\nu &= 4(A_1 A_2), \\ \gamma_\nu \hat{A}_1 \hat{A}_2 \hat{A}_3 \gamma_\nu &= -2\hat{A}_3 \hat{A}_2 \hat{A}_1. \end{aligned} \quad (127.29)$$

Making use of these expressions it is easy to carry out the summation over μ and ν in (127.27).

Thus setting $F_1 = F'_1 + F''_1$ we find

$$\begin{aligned} \text{Tr } F'_1 &= \text{Tr} [\gamma_\nu \omega_2^{-1} (i\hat{q}_2 - m) \gamma_\mu (i\hat{p}_1 - m) \gamma_\mu \omega_2^{-1} (i\hat{q}_2 - m) \gamma_\nu (i\hat{p}_2 - m)] = \\ &= \text{Tr} [\omega_2^{-1} (i\hat{q}_2 - m) \gamma_\mu (i\hat{p}_1 - m) \gamma_\mu \omega_2^{-1} (i\hat{q}_2 - m) \gamma_\nu (i\hat{p}_2 - m) \gamma_\nu] = \\ &= 4\omega_2^{-2} \text{Tr} [(i\hat{q}_2 - m)(i\hat{p}_1 + 2m)(i\hat{q}_2 - m)(i\hat{p}_2 + 2m)], \end{aligned} \quad (127.30)$$

$$\begin{aligned} \text{Tr } F''_1 &= -\text{Tr} [\gamma_\mu \omega_1^{-1} (i\hat{q}_1 - m) \gamma_\nu (i\hat{p}_1 - m) \gamma_\mu \omega_2^{-1} (i\hat{q}_2 - m) \gamma_\nu (i\hat{p}_2 - m)] = \\ &= -4\omega_1^{-1} \omega_2^{-1} \text{Tr} \{ [2i(qq_2)\hat{p}_1 + m\hat{q}_2\hat{p}_1 - im\hat{q}_2(i\hat{q}_1 - m) - \\ &\quad - im\hat{p}_1(i\hat{q}_1 - m) + m^2(i\hat{q}_1 + 2m)](i\hat{p}_2 - m) \}. \end{aligned} \quad (127.31)$$

In calculating the traces it is necessary to use the following rules:

* See, for example, A.I. Akhiezer and V.B. Berestetskii, *Quantum electrodynamics* (Interscience Publ., New York, 1965).

- (1) the trace of the product of an odd number of vectors \hat{A} is equal to zero;
 (2) the trace of a scalar quantity is equal to its quadrupled value;

$$(3) \quad \text{Tr } \hat{A}_1 \hat{A}_2 = 4(A_1 A_2);$$

$$(4) \quad \text{Tr } \hat{A}_1 \hat{A}_2 \hat{A}_3 \hat{A}_4 = 4[(A_1 A_2)(A_3 A_4) + (A_1 A_4)(A_2 A_3) - (A_1 A_3)(A_2 A_4)]. \quad (127.32)$$

The first two rules are trivial. Rules (3) and (4) are easily proved, making use of the identity

$$\hat{A}_1 \hat{A}_2 + \hat{A}_2 \hat{A}_1 = 2(A_1 A_2). \quad (127.33)$$

If we take the trace of the left-hand and right-hand sides and carry out the cyclic permutation of the vectors \hat{A}_1 and \hat{A}_2 under the sign of trace, then we immediately obtain rule (3). We find the trace of $\hat{A}_1 \hat{A}_2 \hat{A}_3 \hat{A}_4$. Making use of the identity (127.33) we have

$$\begin{aligned} \text{Tr } \hat{A}_1 \hat{A}_2 \hat{A}_3 \hat{A}_4 &= -\text{Tr } \hat{A}_1 \hat{A}_2 \hat{A}_4 \hat{A}_3 + 2 \text{Tr } \hat{A}_1 \hat{A}_2 (A_3 A_4) = \\ &= \text{Tr } \hat{A}_1 \hat{A}_4 \hat{A}_2 \hat{A}_3 - 2 \text{Tr } (A_2 A_4) \hat{A}_1 \hat{A}_3 + 8(A_1 A_2)(A_3 A_4) = \\ &= -\text{Tr } \hat{A}_4 \hat{A}_1 \hat{A}_2 \hat{A}_3 + 2 \text{Tr } \hat{A}_2 \hat{A}_3 (A_1 A_4) - 8(A_2 A_4)(A_1 A_3) + 8(A_1 A_2)(A_3 A_4) = \\ &= -\text{Tr } \hat{A}_1 \hat{A}_2 \hat{A}_3 \hat{A}_4 + 8(A_2 A_3)(A_1 A_4) - 8(A_2 A_4)(A_1 A_3) + 8(A_1 A_2)(A_3 A_4). \end{aligned}$$

From this equality we obtain rule (4).

By means of the above rules we find

$$\begin{aligned} \text{Tr } F'_1 &= \frac{16}{\omega_2^2} [2(q_2 p_1)(q_2 p_2) - (q_2^2 + m^2)(p_1 p_2) + 4m^2(q_2 p_1) - \\ &\quad - 4m^2 q_2^2 + 4m^2(q_2 p_2) + 4m^4], \\ \text{Tr } F''_1 &= \frac{16}{\omega_1 \omega_2} [2(q_1 q_2)(p_1 p_2) + m^2(q_2 p_1) + m^2(q_2 q_1) + m^2(q_2 p_2) + \\ &\quad + m^2(q_1 p_1) + m^2(p_1 p_2) + m^2(q_1 p_2) + 2m^4]. \quad (127.34) \end{aligned}$$

Carrying out the replacement mentioned, we obtain from these expressions $\text{Tr } F_2$.

Performing the necessary transformations, after several long but simple calculations, we arrive at the well-known Klein–Nishina formula

$$d\sigma = \frac{1}{2} r_0^2 \left(\frac{\omega_2}{\omega_1} \right)^2 \left(\frac{\omega_1}{\omega_2} + \frac{\omega_2}{\omega_1} - \sin^2 \vartheta \right) d\Omega, \quad (127.35)$$

which plays an important role in applications.

For low photon energies $\omega_1 \ll m$, $\omega_2 = \omega_1$ (see (127.6)) and formula (127.35) in the limit reduces to the classical Thomson formula

$$d\sigma = \frac{1}{2} r_0^2 (1 + \cos^2 \vartheta) d\Omega,$$

obtained in §36 of Part I.

§128. The shift of the terms of the hydrogen atom under the action of the vacuum field (the Lamb shift)

The importance of Feynman's method of calculation does not, of course, reduce to a simplification and standardization of calculations.

As we have already pointed out in §124, the Feynman formalism made it possible to obtain in an obvious form the solution of a number of important problems of quantum electrodynamics. They include, in particular, the Lamb shift of atomic terms already mentioned.

The phenomenon of the Lamb shift yields a very obvious illustration of the validity of those concepts which were assumed as the basis of the quantum theory of radiation and the theory of the positron. In the quantum theory of radiation it was assumed that in a vacuum there is an electromagnetic field. This is the field which corresponds to the zero-point oscillations of field oscillators. The set of electromagnetic field oscillators in states with zero energy is often said to represent the 'electromagnetic vacuum'. In the electromagnetic vacuum, corresponding to the field state with the lowest energy, there is a field strength different from zero. More precisely, the mean (over time) values of the squares of field strengths $(\mathcal{E})^2$ and $(\mathcal{H})^2$ are different from zero.

The existence of the vacuum field had no effect on the phenomena of emission, absorption and scattering which were considered in ch. 12. All these phenomena were associated with the transitions of field oscillators from non-excited (zero) states into excited states and vice versa. Hence in the course of a number of years the properties of the electromagnetic vacuum were not related to directly observed phenomena.

In the theory of positrons it is assumed that in addition to the electromagnetic vacuum there is an electron-positron vacuum, or a background of occupied states with negative energies, which was considered in detail in §116. It turns out that the existence of the electromagnetic and electron-positron vacua is shown directly not only in processes occurring at large energies (for example, in the Compton effect or in the process of pair production) but also in features of the behaviour of particles at small energies, in

particular in the phenomenon of the Lamb shift. The phenomenon of the Lamb shift can be studied rigorously by means of the Feynman formalism. It turns out, however, that this effect can also be discussed without using a relatively complicated mathematical apparatus, on the basis of simple and direct considerations*.

For this we shall first of all discuss the problem as to what magnitude the mean square value of the field strength can have at an arbitrary point of vacuum.

To calculate the mean square value of the field strength in vacuum we consider the normalization volume V_0 . The zero-point oscillation frequency ω has energy $\frac{1}{2}\hbar\omega$. One can write the obvious equality

$$\frac{1}{2}\hbar\omega = \frac{1}{8\pi} \int (\mathcal{E}_{0\omega}^2 + \mathcal{A}_{0\omega}^2) dV = \frac{1}{4\pi} \int \mathcal{E}_{0\omega}^2 dV = \frac{\mathcal{E}_{0\omega}^2 V_0}{8\pi}, \quad (128.1)$$

where $\mathcal{E}_{0\omega}$ and $\mathcal{A}_{0\omega}$ are the field strength amplitudes of the field in vacuum corresponding to zero-point oscillations with frequency ω ; the bar denotes averaging over the oscillation period. Equality (128.1) makes it possible to find the mean square value of the amplitude of the zero-point oscillations of the field with frequency ω :

$$\mathcal{E}_{0\omega}^2 = \frac{4\pi\hbar\omega}{V_0}. \quad (128.2)$$

Let us consider the electron in the hydrogen atom. This electron is acted upon by the Coulomb field of the nucleus and by the fluctuations of the zero point field of the vacuum. Hence a random motion under the action of the vacuum field will be superposed on the orbital motion of the electron.

Let $U(\mathbf{r})$ denote the potential energy of the electron at point \mathbf{r} . We now assume that the coordinate of the electron can be written as $\mathbf{r} = \mathbf{r}_0 + \mathbf{r}'$, where \mathbf{r}_0 is the ordinary value of the coordinate of the electron smoothly varying in its orbital motion, and \mathbf{r}' is its small displacement under the action of a random force, the fluctuating field. Then the change in the mean potential energy of the electron undergoing random displacements can be written in the form

$$\begin{aligned} \langle \Delta U \rangle &= \langle U(\mathbf{r}_0 + \mathbf{r}') - U(\mathbf{r}_0) \rangle \approx \left\langle x'_i \frac{\partial U}{\partial x_i} + \frac{1}{2} (x'_i x'_k) \frac{\partial^2 U}{\partial x_i \partial x_k} \right\rangle = \\ &= \frac{1}{2} \nabla^2 U \langle (x'_i)^2 \rangle = \frac{1}{6} (\nabla^2 U) \langle (\mathbf{r}')^2 \rangle. \end{aligned} \quad (128.3)$$

* See T. Welton, Phys. Rev. 74 (1948) 1157.

Here the bracket $\langle \rangle$ denotes the mean over all possible values of the random quantity \mathbf{r}' . In averaging we have taken into account that $\langle x'_i \rangle = 0$ and that by virtue of the spatial isotropy of random displacements

$$\langle x'_i x'_k \rangle = \frac{1}{3} \langle (\mathbf{r}')^2 \rangle.$$

The value of the potential energy in the factor $\nabla^2 U$ is evidently taken at the value $\mathbf{r} = \mathbf{r}_0$.

The potential energy of the electron in the atom without the perturbation caused by the vacuum field does not depend on the state of the vacuum field and the sign in it is dropped.

In the Coulomb field of the proton one can write for $\nabla^2 U(\mathbf{r}_0)$

$$\nabla^2 U(\mathbf{r}_0) = 4\pi e^2 \delta(\mathbf{r}_0),$$

so that

$$\langle U \rangle = U(\mathbf{r}_0) + \frac{2}{3} \pi e^2 \delta(\mathbf{r}_0) \langle \mathbf{r}'^2 \rangle. \quad (128.4)$$

To obtain the shift of the atomic term we have to take the mean value of (128.4) over the state of the electron in the atom. We then have

$$\Delta E_{\text{Lamb}} = \overline{\langle U \rangle} - \overline{U(\mathbf{r}_0)} = \overline{\langle \Delta U \rangle} = \frac{2}{3} \pi e^2 \int \delta(\mathbf{r}_0) |\psi_n(\mathbf{r}_0)|^2 \langle \mathbf{r}'^2 \rangle dV_0, \quad (128.5)$$

where ψ_n is the wave function of the electron in the atom. Making use of the properties of the δ -function, we find

$$\Delta E_{\text{Lamb}} = \overline{\langle \Delta U \rangle} = \frac{2}{3} \pi e^2 |\psi_n(0)|^2 \langle \mathbf{r}'^2 \rangle. \quad (128.6)$$

The calculation of $\langle \mathbf{r}'^2 \rangle$, the mean square displacement of the electron under the action of the zero-point oscillations of the field, can be carried out relatively simply if only relatively low oscillation frequencies of the field are taken into account.

We shall assume that the displacement of the electron under the action of the field proceeds independently of the orbital motion. Neglecting relativistic effects, one can write the equations of motion in the form

$$m \frac{d^2 \mathbf{r}'_\omega}{dt^2} = e \mathcal{E}_{0\omega} \sin(\mathbf{k} \cdot \mathbf{r} - \omega t),$$

whence

$$\mathbf{r}'_\omega = -\frac{e \mathcal{E}_{0\omega}}{m \omega^2} \sin(\mathbf{k} \cdot \mathbf{r} - \omega t)$$

and, correspondingly,

$$\overline{\langle (r'_\omega)^2 \rangle} = \frac{e^2}{2m^2\omega^4} \mathcal{E}_{0\omega}^2 = \frac{2\pi e^2 \hbar}{m^2\omega^3 V_0}, \quad (128.7)$$

where the bar denotes average over time. Here r'_ω denotes the displacement under the action of the zero-point oscillations of the field with frequency ω .

Since zero-point oscillations with different frequencies are independent, their contribution to the total mean square displacement of the electron is found by simple summation. Consequently, we can write for the total mean square displacement the expression

$$\langle r'^2 \rangle = \int \overline{\langle (r'_\omega)^2 \rangle} \frac{\omega^2 d\omega V_0}{\pi^2 c^3} = \frac{2e^2 \hbar}{\pi c^3 m^2} \int_{\omega_{\min}}^{\omega_{\max}} \frac{d\omega}{\omega}, \quad (128.7')$$

where the integration is carried out over all possible frequencies of zero-point oscillations.

If there were no electron-positron vacuum, then the frequencies of the zero-point oscillations of the field could take on values as large as one wished and the formula obtained would not make sense.

It turns out, however, that at frequencies larger than the minimum frequency of pair production $\omega_1 = 2mc^2/\hbar$ an interaction arises between the zero-point oscillations of the field and the occupied background of negative energies (electromagnetic and electron-positron vacuums). This interaction can be pictured in an obvious way as the interaction between the fluctuations of the 'current' associated with the random displacement of the electron with positive energy and the fluctuations of 'currents' associated with random displacements of the electrons of the background of occupied states, caused by the action of the zero-point oscillations of the electromagnetic vacuum. Since by virtue of the Pauli principle all electrons tend to avoid each other (see §67), it turns out that the fluctuations of the electrons of the background are in the opposite phase with respect to the fluctuations of the electron with positive energy. As a result they are mutually cancelled and the mean square displacement of the electron turns out to be considerably smaller than that given by formula (128.7). Simplifying the actual state of affairs, it can be said that at frequencies larger than ω_1 the mean square displacement of the electron reduces to zero.

Hence as the upper limit of integration over the frequencies of the field in (128.7') one has to choose the quantity $\omega_{\max} = \omega_1$. In defining the minimum frequency ω_{\min} it is necessary to take into account that the electron considered is not free but is bound in an atom.

The frequency ω_{\min} is in order of magnitude equal to the Rydberg frequency of the electron in the hydrogen atom, i.e.

$$\omega_{\min} = \omega_0 = 2\pi R = \frac{me^4}{2\hbar^3},$$

where R is the Rydberg constant equal to $|E_0|/2\pi\hbar$, and E_0 is the energy of the ground state of the atom. Making use of these expressions for ω_{\min} and ω_{\max} we find

$$\langle \mathbf{r}'^2 \rangle = \frac{2e^2\hbar}{\pi c^3 m^2} \ln \frac{2mc^2}{\hbar\omega_0} = \frac{2e^2}{\pi\hbar c} \left(\frac{\hbar}{mc} \right)^2 \ln \frac{2mc^2}{\hbar\omega_0} \quad (128.8)$$

and for the shift of the term we find

$$\Delta E_{\text{Lamb}} = \frac{4e^4}{3\hbar c} \left(\frac{\hbar}{mc} \right)^2 |\psi_n(0)|^2 \ln \frac{2mc^2}{\hbar\omega_0}. \quad (128.9)$$

Formula (128.9) shows that the shift of the levels of the electron in the hydrogen atom under the action of the vacuum on the electron occurs only in s-states. Indeed, only in s-states is the quantity $|\psi_n(0)|^2$ different from zero. This shift is always positive: the level in an s-state must lie above that defined by formula (119.2).

The calculation of ΔE_{Lamb} is of an absolute character, and the numerical value of the shift (with certain additional corrections which are not taken into account in the simplified derivation presented above) is equal to 1057.19 Mc. The experimental value of this quantity turned out to be equal to 1057.77 ± 0.1 Mc. The perfect agreement between the calculated and measured values of the Lamb shift is an obvious confirmation of the general concepts of the reality of the 'vacuum'.

Analogous calculations, on which we shall not dwell, made it possible to find the correction to the magnetic moment of the electron already mentioned (see the article of Welton cited earlier). Of particular importance was the solution in quantum electrodynamics of a number of problems of principle. One succeeded in constructing a quantitative theory which made it possible to calculate with any degree of accuracy the probabilities of all possible processes associated with the interaction of electrons with each other and with the electromagnetic field.

The difficulties of principle of the theory associated with diverging expressions, for example the difficulty frequently mentioned that the intrinsic mass (or energy) of the electron goes to infinity, could, to a certain degree, be removed. In expressions for the intrinsic energy of a particle and in similar relations one succeeded in separating finite observable quantities, whereas

diverging expressions describe only quantities which are in principle not observable.

This procedure, called renormalization, cannot be presented here, and we refer the reader to the specialist literature, for example to the monographs of S.Schweber, H.Bethe and F.de Hoffman or A.I.Akhiezer and V.B.Berestetskii, which have frequently been cited before.

Fundamentals of the Theory of Elementary Particles

§129. The classification and properties of elementary particles

At the present time a large number (of the order of two hundred) of elementary particles have been discovered which can transform into each other but which do not consist, in the usual sense of the word, of smaller entities. They can be divided into two large groups: stable particles and short-lived particles (resonances). The first term is a relative one, since the group of stable particles contains, for example, both the electron which lives infinitely long, and the π^0 -meson, whose lifetime is of the order of 10^{-16} sec. Stability is understood only in the sense that the lifetime of these particles is much longer than the characteristic time of 10^{-24} – 10^{-23} sec for a light signal to traverse a distance of 10^{-13} cm which is typical of the 'size' of elementary particles.

The stable particles are divided into the following four classes:

1. The class of photons, which comprises the quanta of classical fields: This class includes the photon itself, a quantum of the electromagnetic field, and sometimes the graviton, a quantum of the gravitational field.
2. The class of leptons (light particles) contains the electron, the muon, which was earlier called the μ -meson, two neutrinos (the electron neutrino and the muon neutrino) and the four corresponding antiparticles. The elec-

iron neutrino is produced in the decay of the neutron, and the muon neutrino in the decay of the muon:

$$n \rightarrow p + e^- + \bar{\nu}_e, \quad \mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$$

(the bar distinguishes antiparticles from the corresponding particles).

3. The class of mesons (particles of medium mass) comprises 3 π -mesons (π^- is the antiparticle of π^+ , while the particle and antiparticle for π^0 are identical), 4 K-mesons (two particles and two antiparticles) and 1 η -meson (which is its own antiparticle).

4. The class of baryons (heavy particles) contains two nucleons (the proton and the neutron), 1 Λ -hyperon, 3 Σ -hyperons, 2 Ξ -hyperons, 1 Ω -hyperon and the corresponding antiparticles.

The term 'resonance' arose in connection with the fact that data on short-lived particles were initially obtained from scattering experiments. Characteristic resonance maxima were observed in total cross sections at definite values of the energy of the scattered particles. Thus, for example, in experiments with π -meson scattering the existence of the ρ -meson was discovered.

The group of short-lived particles comprises only baryon resonances and meson resonances — in all more than 150 particles. The most important among them are the mesons ω (1), φ (1), ρ (3), K^* (4) and the baryon resonances Δ_{1236} (4), Σ_{1385} (3), Ξ_{1530} (2) along with their antiparticles. Furthermore, the f -meson, which was theoretically predicted by Pomeranchuk and subsequently discovered experimentally, is of interest.

The set of all mesons and baryons and their resonances forms a large group of particles which are at present called hadrons. Also the term hadenons, referring to leptons and to the photon, has begun to be used in Russian particle physics literature.

Let us briefly enumerate the basic properties of elementary particles.

1. Each particle possesses a rest mass, which is measured in MeV. The range of mass values of different particles is rather wide: from 0 (photon and neutrino) up to 3000 MeV and more (Δ_{3230}). The initial classification of particles (their division into the four classes mentioned above) was based only on the values of their masses. But it turned out that this classification, just as with the mass of atoms in the periodic system of elements, is rather loose. This is particularly clearly seen in examples of resonances.

2. A very important characteristic of a particle is the value of its spin σ . The photon has spin 1 (with certain reservations, since its rest mass is equal to zero). Leptons have spin $\frac{1}{2}$, stable mesons spin 0, baryons (except for Ω) spin $\frac{1}{2}$, and the Ω -hyperon spin $\frac{3}{2}$. Among resonances there are particles with spin values from 0 up to $\frac{19}{2}$. The spin of the meson resonances listed above

(except for f) is equal to 1, the spin of the baryon resonances is equal to $\frac{3}{2}$, and that of the f -meson is 2. According to the Pauli—Lüders theorem, proved on the basis of most general principles which do not depend on actual dynamics, the spin of a particle unambiguously defines the type of statistics: particles with half-integer spin (leptons, baryons and baryon resonances) obey Fermi—Dirac statistics, while those with integer spin (photon, mesons and meson resonances) obey Bose—Einstein statistics. Furthermore, the spin of the particle determines the transformation properties of its wave function with respect to Lorentz transformations. Spin zero particles are described by a (pseudo-)scalar wave function, spin one-half particles by a spinor wave function, spin one-half particles by a (pseudo-)vector wave function and so on.

3. The parity P of a particle determines the transformation properties of its wave function with respect to the space inversion transformation. All stable mesons have odd parity and, having zero spin, are described by a pseudoscalar wave function. Meson resonances of spin 1 ($\omega, \varphi, \rho, K^*$) also have odd parity. Since under space inversion the vector components change sign, the wave function of these particles is vectorial. The parity of the f -meson is even. Baryons and their resonances can be assigned only relative parity. If it is assumed by definition that the parities of the proton, neutron and Λ -hyperon are even, then the parities of all the baryons enumerated above and of their resonances will also be even, whereas those of the antibaryons will be odd.

4. Each particle is characterized by the value of its electric charge, which (the charge of the electron being assumed to be equal to unity) can take on only integer values. At present a large number of neutral particles and particles with a charge equal in absolute value to unity are known. The charge of six particles (Δ -resonances) is equal to +2.

5. In order to characterize elementary particles, the lepton number L and baryon number B are introduced. By definition, for leptons $L = +1$, $B = 0$; for antileptons $L = -1$, $B = 0$; for baryons $L = 0$, $B = +1$, for antibaryons $L = 0$, $B = -1$; for mesons and photons $L = B = 0$. In all reactions involving elementary particles these quantum numbers are conserved, and their importance is due just to this fact.

6. All hadrons are divided into small families whose members are denoted by one and the same symbol (for example π). These families are called isomultiplets. The particles constituting an isomultiplet have about the same mass but different charges. Each isomultiplet is assigned a definite value of isospin T , which defines the number of members of the multiplet $N = 2T + 1$. Thus the isospin of the nucleon and of the K -meson is equal to $\frac{1}{2}$, the isospin of the Σ -hyperon and π -meson is equal to 1, that of the Δ -resonance is equal to $\frac{3}{2}$ and so on.

Table 3

Stable elementary particles

Class	Particle	Mass (MeV)	Spin and parity	Lifetime (sec)	Basic decays	S	T	T_3
Photon $B, L=0$	γ	0	1^-	∞	—	—	—	—
Leptons $B=0, L=1$	ν_e	0	$\frac{1}{2}^+$	∞	—	—	—	—
	ν_μ	0	$\frac{1}{2}^+$	∞	—	—	—	—
	e^-	0.511	$\frac{1}{2}^+$	∞	—	—	—	—
	μ^-	105.66	$\frac{1}{2}^+$	2.2×10^{-6}	$e^- \bar{\nu}_e \nu_\mu$	—	—	—
Mesons $B=0, L=0$	π^+	139.58	0^-	2.6×10^{-8}	$\mu^+ \nu_\mu$	0	1	$+1$
	π^0	134.98		0.89×10^{-16}	$\gamma\gamma$			0
	π^-	139.58		2.6×10^{-8}	$\mu^- \bar{\nu}_\mu$			-1
	K^+	493.8	0^-	1.24×10^{-8}	$\mu^+ \nu_\mu; \pi^+ \pi^0; \pi^+ \pi^- \pi^+$	$+1$	$\frac{1}{2}$	$+\frac{1}{2}$
	$K^0 \left(\begin{smallmatrix} K_S^0 \\ K_L^0 \end{smallmatrix} \right)$	497.9		$\left\{ \begin{smallmatrix} 0.87 \times 10^{-10} \\ 5.73 \times 10^{-8} \end{smallmatrix} \right.$	$\pi^+ \pi^-; \pi^0 \pi^0$			$-\frac{1}{2}$
					$\pi\pi\nu_e; \pi\mu\nu_\mu; \pi^0\pi^0\pi^0; \pi^+\pi^-\pi^0$			
Baryons $B=1, L=0$	η	548	0^-	$\sim 10^{-17}$	$\gamma\gamma; \pi^+ \pi^- \pi^0; \pi^0 \pi^0 \pi^0; \pi^0 \gamma\gamma$	0	0	0
	p	938.25	$\frac{1}{2}^+$	∞	—	0	$\frac{1}{2}$	$+\frac{1}{2}$
	n	939.55		$\sim 10^3$	$p e^- \bar{\nu}_e$			$-\frac{1}{2}$
	Λ	1115.6	$\frac{1}{2}^+$	2.54×10^{-10}	$p\pi^-; n\pi^0$	-1	0	0
	Σ^+	1189.5	$\frac{1}{2}^+$	0.8×10^{-10}	$p\pi^0; n\pi^+$	-1	1	$+1$
	Σ^0	1192.6		$< 1 \times 10^{-14}$	$\Lambda\gamma$			0
	Σ^-	1197.4		1.65×10^{-10}	$n\pi^-$			-1
	Ξ^0	1314.7	$\frac{1}{2}^+$	3×10^{-10}	$\Lambda\pi^0$	-2	$\frac{1}{2}$	$+\frac{1}{2}$
	Ξ^-	1321.2		1.74×10^{-10}	$\Lambda\pi^-$			$-\frac{1}{2}$
	Ω	1674	$\frac{3}{2}^+$	$\sim 1 \times 10^{-10}$	$\Xi\pi; \Lambda\bar{K}^+$	-3	0	0

Table 4
The most important resonances

Class	Particle	Mass (MeV)	Spin and parity	Width ^a (MeV)	Basic decays	<i>S</i>	<i>T</i>
Meson resonances $B=0, L=0$	ω	783	1^-	12	$\pi^+\pi^-\pi^0; \pi^0\gamma$	0	0
	φ	1019	1^-	4	$K^+K^-; K_L K_S; \pi^+\pi^-\pi^0$	0	0
	f	1250	2^+	110	$\pi\pi$	0	0
	ρ^\pm	774	$\{1^-$	$\}128$	$\}\pi\pi$	$\}0$	$\}1$
	ρ^0	780					
	K^*	892	1^-	50	$K\pi$	+1	$\frac{1}{2}$
Baryon resonances $B=1, L=0$	Δ_{1236}	1236	$\frac{3}{2}^+$	120	$N\pi$	0	$\frac{3}{2}$
	Σ_{1385}	1382	$\frac{3}{2}^+$	40	$\Lambda\pi; \Sigma\pi$	-1	1
	Ξ_{1530}	1530	$\frac{3}{2}^+$	7	$\Xi\pi$	-2	$\frac{1}{2}$

^a The lifetime τ is connected with the width Γ by the uncertainty relation $\Delta E \Delta \sim \hbar$, i.e. $\tau = \hbar/\Gamma$.

7. The different particles constituting an isomultiplet differ from each other by the isospin projection T_3 onto the third axis of the fictitious isospace. Depending on the value of isospin T , its projection T_3 can take on integer or half integer values. The concept of isospin was initially introduced only for the nucleon and pion. In this case T_3 is related in the following way to the value of the electric charge of the particle:

$$Q = T_3 + \frac{1}{2}B. \quad (129.1)$$

8. After the discovery of the K-mesons and hyperons (the so-called strange particles) it was necessary to modify the above formula:

$$Q = T_3 + \frac{1}{2}(B+S) \quad (129.2)$$

(the Gell-Mann—Nishijima relation). The new quantum number S was called strangeness. In a wide range of phenomena, for example in hadron production reactions, the strangeness is conserved, which immediately made it possible to explain certain incomprehensible features of these processes (say, the fact that strange particles are always produced in pairs). Recently instead of strangeness the physicist prefers to use another quantum number, the hypercharge Y , which is closely related to S :

$$Y = B + S. \quad (129.3)$$

We shall not dwell on some other quantum numbers, which are introduced to characterize elementary particles (time parity and charge parity, G -parity, muonic charge and so on).

A summary of properties is given in tables 3 and 4.

§ 130. The types of interactions of elementary particles

Elementary particles can take part in very different types of interaction: a particle may annihilate with its antiparticle, fast particles in collisions are scattered and new particles are produced, many particles are unstable and disintegrate and so on. At present four types of elementary particle interaction, sharply differing from each other in strength and other properties, are known.

1. The electromagnetic interaction: the interaction of charged particles with photons, and by means of them also with each other. Because of virtual processes neutral particles may also take part in the electromagnetic interaction. Examples of reactions caused by the electromagnetic interaction are the transformations

$$e^- + e^+ \rightarrow 2\gamma, \quad \gamma + e^- \rightarrow \gamma + e^- \quad (\text{Compton effect}),$$

$$\pi^0 \rightarrow 2\gamma, \quad \Sigma^0 \rightarrow \Lambda + \gamma,$$

and so on. The electromagnetic interaction is of infinite range; times of 10^{-16} – 10^{-14} sec are characteristic for reactions caused by it. The strength of the electromagnetic interaction is determined by the charge of the particle or by the dimensionless coupling constant which, in this particular case, is the fine structure constant $\alpha = e^2/\hbar c = 1/137$. The smallness of this constant allows the electromagnetic interaction to be considered as a perturbation, which explains the successes of quantum electrodynamics (see ch. 14).

2. The strong interaction: the interaction of hadrons which is responsible for their scattering, for production reactions and for resonance decays. Typical examples are

$$\pi + N \rightarrow \pi + N, \quad \pi^- + p \rightarrow \Sigma^- + K^+,$$

$$\Delta \rightarrow N + \pi, \quad \rho \rightarrow 2\pi,$$

and so on. The strong interaction is of short range, the radius of action being of the order of 10^{-13} cm; times of 10^{-24} – 10^{-23} sec are characteristic of it. The intensity of the strong interaction is characterized by a parameter g , which is an analogue of the electric charge. For the interaction of π -mesons with nucleons the dimensionless coupling constant is equal to $g^2/\hbar c \approx 14$, so that the strong interaction is three orders of magnitude greater than the electromagnetic interaction. Therefore perturbation theory is inadequate for its analysis. Up to now there is no complete theory of strong interactions.

3. The weak interaction: this is responsible for the slow decays of elementary particles, for example

$$p \rightarrow n + e^+ + \bar{\nu}_e, \quad \mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu,$$

$$K^+ \rightarrow \pi^+ + \pi^0 \quad (\theta\text{-decay}), \quad K^+ \rightarrow \pi^+ + \pi^- + \pi^+ \quad (\tau\text{-decay}).$$

It can in principle also give rise to other reactions (for example, the scattering of neutrinos by electrons), but such processes have extremely small cross sections and have not been observed experimentally. The weak interaction is of an even shorter range than the strong interaction (its range may be of the order of 10^{-17} cm), and its characteristic times are 10^{-10} – 10^{-6} sec. To characterize the intensity of the weak interaction one cannot introduce a natural dimensionless constant as in the preceding cases. This is associated with the fact that the analogue of electric charge, the so-called Fermi constant G , has a dimensionality different from that of e . The dimensionless constant of the weak interaction must involve a certain mass. If the mass of

the π -meson is taken, then $G^2(\hbar c)^{-2}(\hbar/\mu c) \approx 5 \times 10^{-14}$, i.e. the weak interaction is less strong than the electromagnetic interaction by about 11 orders of magnitude. Nevertheless, strictly speaking, perturbation theory cannot be used in this case, because the weak interaction is non-renormalizable (see ch. 14).

4. The gravitational interaction: this is characterized by an extremely small dimensionless coupling constant $\kappa M/\hbar c = 2 \times 10^{-39}$ (κ is the gravitational constant, and M is the mass of the nucleon), which allows one to disregard it at the present state of development of the theory of elementary particles.

The interactions enumerated above differ not only in strength but also in their conservation laws, which is, apparently, even more essential. In all the interactions the energy, momentum, angular momentum, electric charge and baryon and lepton numbers are conserved. However, there are no universal conservation laws for isospin T , its projection T_3 , strangeness S (or hypercharge Y) and parity P .

1. The strong interaction is the most symmetric; in the reactions to which it gives rise all the quantum numbers mentioned above are conserved. In particular, the isospin conservation law is an expression of charge independence: all the terms of an isomultiplet behave in the same way with respect to strong interactions.

2. 'Switching on' the electromagnetic interaction violates the equivalence of the particles contained in an isomultiplet, since they have different charges. Apparently, it is just this interaction which is responsible for the existence of the small differences between the masses of these particles. Since switching on the electromagnetic interaction specifies a definite direction in isospace, the total isospin T will no longer be conserved, but its projection T_3 is still conserved. Also the laws of conservation of strangeness S and of parity P hold.

3. The weak interaction is the least symmetric. None of the four quantum numbers mentioned above is conserved in the corresponding reactions. In particular, the parity conservation law is no longer valid (see ch. 13). Moreover, as shown by experiments with neutral K-mesons, the weak interaction is apparently also not invariant under time reversal.

§131. Symmetry groups in quantum mechanics

We have already stressed that no consistent and complete theory of strong interactions has been developed up to now. In particular, dynamical equations describing the behaviour of particles under the strong interaction have not

been formulated. Therefore the study of the general symmetry properties of strong interactions assumes a special role. It allows one to obtain a satisfactory classification of hadrons and to derive a number of quantitative relations.

In §129 it was pointed out that all hadrons are divided into small families, isomultiplets, which can be assigned definite values of the isospin T . The members of a given multiplet differ in the isospin projection T_3 , which determines the value of electric charge, and when the electromagnetic interaction is 'switched off' they have strictly the same mass. In strong interactions the quantum numbers T and T_3 are conserved.

We have frequently encountered analogous situations already. For example, when a non-relativistic particle moves in a central field its possible states are also grouped into definite sets which are characterized by different values of the angular momentum J . The wave functions belonging to the same set differ in the angular momentum component J_3 and corresponds to one and the same energy, i.e. form a degenerate energy level. In the motion of the particle the quantum numbers J and J_3 are conserved.

The invariance of theory with respect to a definite class of transformations in real or in a certain fictitious space (in our quantum-mechanical example with respect to space rotations) is characteristic of all similar cases. The set of transformations is closed, i.e. successive application of the allowed transformations again leads to an allowed transformation. Furthermore, there is an identity or unit transformation, and to each transformation there corresponds an inverse transformation. Such invariance (symmetry) transformations are said to form a group. Its elements can be denoted by g , and successive application of two transformations g_1 and g_2 will be written in the form of a product $g_2 g_1$ (in just this order). It is easily seen that all space rotations form a group — a three-dimensional rotation group, denoted by $O(3)$. Lorentz transformations are another example of transformations forming a group.

The concept of the invariance of a theory with respect to a given group of transformations involves two aspects: definite transformation properties of the wave functions ψ and definite transformation properties of the Hamiltonian \hat{H} .

With respect to a group of transformations g the entire Hilbert space of the wave functions is broken into invariant sub-spaces, i.e. there are sets of wave functions which transform, according to a given law, only into each other:

$$\psi' = U(g)\psi . \quad (131.1)$$

It is required that to the product $g_2 g_1$ of the elements of the group there correspond the product of the operators $U(g)$:

$$U(g_2 g_1) = U(g_2)U(g_1). \quad (131.2)$$

In this case the set of operators $U(g)$ is said to form a representation of the given group. The dimensionality of the space (the maximum number of linearly independent wave functions) in which these operators act is said to be the dimension of the given representation. If the invariant sub-space does not contain any invariant sub-spaces of lower dimension, then one speaks of an irreducible representation. Otherwise the representation is said to be reducible. In what follows the set of wave functions transformable according to an irreducible representation of a symmetry group will be called a multiplet.

From the theory of angular momentum it is known that in space rotations the spherical functions $Y_J^m(\theta, \varphi)$ corresponding to a given angular momentum J and to all its possible projections J_z transform only into each other. There corresponds to them an irreducible representation with dimension $2J+1$ of the group $O(3)$, i.e. all spherical functions corresponding to the angular momentum J form a $(2J+1)$ -dimensional multiplet of the rotation group.

Let us consider the requirements which, in the invariant theory, are imposed upon the transformation properties of the Hamiltonian. We act on the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H} \psi \quad (131.3)$$

with the operator $U(g)$ of the representation according to which the wave function ψ transforms. Assuming that $U(g)$ commutes with the operator $\partial/\partial t$ (we shall not need the analysis of the more general case), we obtain

$$i\hbar \frac{\partial (U\psi)}{\partial t} = U \hat{H} U^{-1} U\psi$$

or

$$i\hbar \frac{\partial \psi'}{\partial t} = U \hat{H} U^{-1} \psi'. \quad (131.4)$$

The invariance of the theory means the identity of the form of the Schrödinger equation for the initial wave function ψ and for the transformed wave function ψ' . Hence

$$\hat{H} = U \hat{H} U^{-1}$$

or

$$[\hat{H}, U(g)] = 0. \quad (131.5)$$

Thus the requirement of invariance of a theory with respect to transformations of a certain group leads to the commutativity of the Hamiltonian with all the representation operators of this group.

We shall consider below only the so-called Lie groups, whose elements are single-valued differentiable functions of a finite number of real parameters. The latter are chosen in such a way that a unit element corresponds to their zero values. Thus for a Lie group:

$$g = g(\alpha_1, \dots, \alpha_n), \quad g(0, \dots, 0) = I. \quad (131.6)$$

The number n of all independent real parameters of the Lie group is said to be its dimension. If the transformation g differs infinitely little from the identity transformation, i.e. if it is infinitesimal, then there correspond to it infinitesimal values of the parameters α_k . In this case, taking into account (131.6) one can write

$$g(\alpha_1, \dots, \alpha_n) \approx I + \sum_{k=1}^n \alpha_k \frac{\partial g(\alpha_1, \dots, \alpha_n)}{\partial \alpha_k} \Big|_{\alpha_1=\dots=\alpha_n=0} \equiv I + i \sum_{k=1}^n \alpha_k l_k. \quad (131.7)$$

The quantities

$$l_k = -i \frac{\partial g(\alpha_1, \dots, \alpha_n)}{\partial \alpha_k} \Big|_{\alpha_1=\dots=\alpha_n=0} \quad (131.8)$$

(the factor $-i$ is introduced for convenience), called group generators, represent square matrices whose order is equal to the dimension of the space in which the group transformations act. The rotation group generators are the 3×3 matrices J_x, J_y, J_z presented in (51.17). To see this it is sufficient to take the matrices of finite rotations about the corresponding coordinate axes and to make use of the definition (131.8).

It is obvious that the operators $U(g)$ of the representations of the Lie group depend on the same parameters α_k . Taking into account that to the unit element of the group there corresponds the unit representation operator, one can write

$$U(g) = U(\alpha_1, \dots, \alpha_n), \quad U(0, \dots, 0) = I. \quad (131.9)$$

The N -dimensional representation operators corresponding to infinitesimal transformations have the form

$$U(g) = I + \sum_{k=1}^n \alpha_k \left. \frac{\partial U(\alpha_1, \dots, \alpha_n)}{\partial \alpha_k} \right|_{\alpha_1 = \dots = \alpha_n = 0} \equiv I + i \sum_{k=1}^n \alpha_k L_k. \quad (131.10)$$

Here the quantities

$$L_k = -i \left. \frac{\partial U(\alpha_1, \dots, \alpha_n)}{\partial \alpha_k} \right|_{\alpha_1 = \dots = \alpha_n = 0} \quad (131.11)$$

called representation generators, are $N \times N$ matrices. For the rotation group such operators are, to within the factor $\frac{1}{2}\hbar$, the operators $\hat{J}_x, \hat{J}_y, \hat{J}_z$ of the angular momentum components. In particular, spinor representation generators are the same as the Pauli matrices (see (61.13)). It can be shown that the operators $U(g)$ of a given representation, which correspond to finite transformations, are of the form

$$U(g) = \exp \left(i \sum_{k=1}^n \alpha_k L_k \right) \quad (131.12)$$

(see, for example, formulae (61.13) and (61.12)).

In §46 we have seen that only in the case of unitary transformations of the wave functions does the physical content of a theory not change, which singles out of the multitude of all representations a very important class of unitary representations for which

$$U^\dagger U = U U^\dagger = I. \quad (131.13)$$

From (131.13) and (131.10) we have

$$I = U^\dagger U \approx \left(I - i \sum_{k=1}^n \alpha_k L_k^\dagger \right) \left(I + i \sum_{k=1}^n \alpha_k L_k \right) \approx I + i \sum_{k=1}^n \alpha_k (L_k - L_k^\dagger).$$

Hence

$$L_k^\dagger = L_k, \quad (131.14)$$

i.e. unitary representation generators are Hermitian matrices. Hence for infinitesimal transformations

$$U(g) \approx I + i \sum_{k=1}^n \alpha_k L_k; \quad U^\dagger(g) = U^{-1}(g) \approx I - i \sum_{k=1}^n \alpha_k L_k. \quad (131.15)$$

We also note that for the Lie groups the relation (131.5), valid in the invariant theory, is evidently equivalent to the condition

$$[\hat{H}, L_k] = 0. \quad (131.16)$$

The importance of studying symmetry groups in quantum mechanics is due to the following facts:

1. There exist Hermitian mutually commuting combinations of the representation generators which also commute with them. They are called invariant operators, and their basic property is that on a given multiplet these operators are simply multiples of the unit operators (the Shur lemma). This means that all the wave functions of one multiplet are eigenfunctions of any invariant operator with one and the same eigenvalue. Thus a natural set of quantum numbers arises equal in number to the invariant operators characterizing the multiplet as a whole. In the rotation group there is one invariant operator $\hat{\mathbf{J}}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2$, so that in this case each multiplet is characterized by the value of the angular momentum J .

2. Among the representation generators there may exist several mutually commuting generators. Their number is determined by the properties of the group and is said to be its rank. The basis functions of a multiplet can be chosen in such a way that they are eigenfunctions of these generators. The corresponding eigenvalues are the quantum numbers classifying the wave functions belonging to the given multiplet. In the rotation group there are no mutually commuting generators, i.e. its rank is equal to 1. Therefore wave functions with a given angular momentum can be assigned only one more quantum number, for example the value of the component J_z .

3. The generators and invariant operators described form a set of Hermitian operators which commute with each other and with the Hamiltonian (see (131.16)). Hence there correspond to them conserved and at the same time measurable physical quantities. A consequence of the theory of invariance with respect to space rotations is the conservation of angular momentum \mathbf{J} and of its component J_z .

4. From the above it follows that if the wave function of the initial state of a system belongs to a certain multiplet, then as a result of a reaction (scattering or decay) the system will make a transition to a new state whose wave function belongs to the same multiplet. This establishes definite selection rules for the reactions.

5. From (131.16) and the Shur lemma it follows that the eigenvalues of the Hamiltonian (the values of the energy or mass of the elementary particles) are the same for the wave functions of a given multiplet. This accounts for the presence of degeneracy and allows one to establish the multiplicity, which is equal to the dimension of the multiplet. In a theory invariant with respect to the rotation group there occurs a degeneracy in J_z , its multiplicity being equal to $2J+1$.

§132. The isogroup SU(2) and its representations

The existence of hadron multiplets with the properties already described suggests that the strong interaction of elementary particles is invariant with respect to a certain group of transformations. It turned out that this latter is the group SU(2), which we shall henceforth call the isogroup. From the mathematical point of view it is closely related (is almost equivalent) to the rotation group O(3).

By the group SU(2) is meant the set of all unitary and unimodular (determinant equal to 1) 2×2 matrices:

$$g^\dagger g = gg^\dagger = I; \quad \det g = 1. \quad (132.1)$$

We can make a useful geometrical interpretation of unitary 2×2 matrices. We take the two-dimensional complex space of vectors x , which we shall write in the form of columns, and introduce the Hermitian conjugate vectors x^\dagger :

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \quad x^\dagger = (x_1^*, x_2^*) \equiv (x^1, x^2). \quad (132.2)$$

In this space we consider the linear transformation

$$x' = gx \quad \text{or} \quad x'_i = g_i^j x_j \quad (132.3)$$

(the superscript numbers the column, while the subscript numbers the row; they run over the values 1, 2, repeated indices implying summation). Second-order unitary matrices correspond to the transformation matrices (132.3), which do not change the quadratic form

$$x^\dagger x = x^i x_i = x_1^* x_1 + x_2^* x_2. \quad (132.4)$$

According to §131 the infinitesimal matrix g can be written as

$$g \approx I + i\epsilon_\alpha \tau_\alpha. \quad (132.5)$$

Here τ_α are the generators of the group SU(2), and ϵ_α are its parameters whose number n (the dimension of the group) is to be defined. The requirement of unitarity of the matrix g leads to the hermiticity of its generators τ_α . Noting that, to terms of the second order of smallness in ϵ_α , for the infinitesimal matrix (132.5)

$$1 = \det g \approx 1 + i\epsilon_\alpha \text{Tr} \tau_\alpha$$

we arrive at the conclusion that the trace of the generators τ_α is equal to zero. Thus they are 2×2 matrices with the properties

$$\tau_{\alpha}^{\dagger} = \tau_{\alpha}, \quad \text{Tr } \tau_{\alpha} = 0. \quad (132.6)$$

These restrictions impose 5 (4 plus 1) conditions upon the 8 real parameters of the complex 2×2 matrix. Hence there are 3 independent matrices with the properties (132.6), which just determines the dimension of the group SU(2). One can take as its generators the Pauli matrices, which are Hermitian and have trace zero:

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (132.7)$$

Among these there are no mutually commuting ones, so that the rank of SU(2) is equal to 1. The sum of the squares of the Pauli matrices commutes with each of them:

$$[\tau^2, \tau_{\alpha}] = 0, \quad \text{where} \quad \tau^2 = \tau_1^2 + \tau_2^2 + \tau_3^2. \quad (132.8)$$

Let us now go on to the construction of the representations of the group SU(2).

1. The trivial representation is the simplest one. Its multiplets are one-dimensional, i.e. contain one wave function φ each, which does not change under the transformation (132.3):

$$\varphi' = \varphi. \quad (132.9)$$

The dimension of this representation is 1, and its generators are the numbers 0. Such a representation is called a scalar representation, and φ is said to be a scalar.

2. Let us consider the two-dimensional space of vectors φ_i , which transform according to the same law as the vectors x :

$$\varphi'_i = g^j_i \varphi_j \quad \text{or} \quad \varphi'_i \approx (I + i\epsilon_{\alpha} \tau_{\alpha})^j_i \varphi_j. \quad (132.10)$$

As a result, two-dimensional multiplets arise whose members transform according to the representation of the same dimension. From (132.10) it follows that its generators are the matrices τ_{α} themselves:

$$(L_{\alpha})^j_i = (\tau_{\alpha})^j_i. \quad (132.11)$$

This representation is called the spinor representation, and the quantities φ_i are said to be spinors (see the formalism presented in ch. 8).

3. Let us take two quantities φ^i which transform in the same way as the components of the vector x^{\dagger} :

$$(\varphi^i)' = (g^{\dagger})^i_j \varphi^j \quad \text{or} \quad (\varphi^i)' \approx (I - i\epsilon_{\alpha} \tau_{\alpha})^i_j \varphi^j. \quad (132.12)$$

As a result we obtain a two-dimensional representation, which is said to be conjugate to the spinor representation. Its generators are the matrices

$$(L_\alpha)_i^j = -(\tau_\alpha)_i^j. \quad (132.13)$$

In fact this representation is equivalent to the spinor representation, i.e. from the quantities φ^i it is possible to form linear combinations which transform according to the law (132.10). But its introduction is very convenient from the formal point of view.

4. All other representations can be constructed from the spinor representation and its conjugate. Let us consider the set of $2(p+q)$ quantities $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ which transform as the product of p spinors and q conjugate spinors, i.e. according to the law

$$(\varphi_{i_1 \dots i_p}^{j_1 \dots j_q})' = g_{i_1}^{i'_1} \dots g_{i_p}^{i'_p} (g^\dagger)_{j'_1}^{j_1} \dots (g^\dagger)_{j'_q}^{j_q} \varphi_{i'_1 \dots i'_p}^{j'_1 \dots j'_q}. \quad (132.14)$$

For the infinitesimal transformation we have

$$(\varphi_{i_1 \dots i_p}^{j_1 \dots j_q})' = (I + i\epsilon_{\alpha_1} \tau_{\alpha_1})_{i_1}^{i'_1} \dots (I - i\epsilon_{\beta_q} \tau_{\beta_q})_{j'_q}^{j_q} \varphi_{i'_1 \dots i'_p}^{j'_1 \dots j'_q} \quad (132.15)$$

so that the generators of this representation have the form

$$\begin{aligned} (L_\alpha)_{i_1 \dots i_p j'_1 \dots j'_q}^{i'_1 \dots i'_p j_1 \dots j_q} &= \sum_{s=1}^p \delta_{i_1}^{i'_1} \dots \delta_{i_{s-1}}^{i'_{s-1}} (\tau_\alpha)_{i_s}^{i'_s} \delta_{i'_{s+1}}^{i_{s+1}} \dots \delta_{i'_p}^{i_p} \delta_{j'_1}^{j_1} \dots \delta_{j'_q}^{j_q} - \\ &- \sum_{s=1}^q \delta_{i_1}^{i'_1} \dots \delta_{i_p}^{i'_p} \delta_{j'_1}^{j_1} \dots \delta_{j'_{s-1}}^{j_{s-1}} (\tau_\alpha)_{j'_s}^{j_s} \delta_{j'_{s+1}}^{j_{s+1}} \dots \delta_{j'_q}^{j_q}. \end{aligned} \quad (132.16)$$

The representation obtained in such a way is said to be the direct product of p spinor representations and q conjugate representations. Symbolically

$$(1,0) \otimes \dots \otimes (1,0) \otimes (0,1) \otimes \dots \otimes (0,1)$$

$p \text{ times} \qquad \qquad \qquad q \text{ times}$

(the notation is obvious).

If the function $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ is symmetric with respect to a certain pair of subscripts (or superscripts), then this property is conserved under the transformation (132.13). Furthermore, the invariance of the quadratic form (132.4) leads to the conservation of the trace with respect to any pair in superscripts and subscripts, i.e. quantities of the type $\varphi_{\dots i \dots}^{\dots i \dots}$. Hence the set

of all functions $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ is divided into individual multiplets whose members transform only into each other, i.e. the representation constructed is reducible. In the given case the condition of irreducibility amounts to the fact that $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ must be symmetric separately with respect to all the subscripts and superscripts, the trace with respect to any pair in the subscripts and superscripts being bound to reduce to zero. It is easy to count the number of independent quantities $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ of such a type, i.e. to determine the dimension of the corresponding irreducible representation. The total number of subscripts (superscripts) which are equal to 1 can vary from 0 up to p (up to q), the other indices being equal to 2. Under the condition of symmetry the order of succession of the indices does not matter, so that the number of different completely symmetric quantities $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ is equal to $(p+1)(q+1)$. The fact that the traces are equal to zero imposes upon them pq conditions, so that in all we have

$$N = (p+1)(q+1) - pq = p + q + 1$$

independent components. The irreducible representation described will be denoted by the symbol (p, q) , and its dimension by $N(p, q)$. We have

$$N(p, q) = (p+q) + 1. \quad (132.17)$$

We stress once more that the superscripts and subscripts are equivalent, so that the irreducible representation is in essence defined by the total number of indices, which is equal to $p+q$.

Let us consider the example (important for what follows) of the direct product of the spinor representation by the conjugate representation, i.e. the representation $(1, 0) \otimes (0, 1)$. Functions transforming according to this representation form a square 2×2 matrix. In correspondence with (132.13)

$$(\varphi_i^j)' = g_i^{i'} (g^\dagger)^{j'}_{j'} \varphi_{i'}^{j'}. \quad (132.18)$$

For infinitesimal transformations (132.15) gives

$$\begin{aligned} (\varphi_i^j)' &\approx (I + i\epsilon_\alpha \tau_\alpha)_i^{i'} (I - i\epsilon_\beta \tau_\beta)^{j'}_{j'} \varphi_{i'}^{j'} \\ &\approx \{\delta_i^{i'} \delta_{j'}^{j'} + i\epsilon_\alpha [(\tau_\alpha)_i^{i'} \delta_{j'}^{j'} - (\tau_\alpha)^{j'}_{j'} \delta_i^{i'}]\} \varphi_{i'}^{j'} \end{aligned} \quad (132.19)$$

so that the generators of this representation have the form

$$(L_\alpha)_{ij}^{i'j'} = (\tau_\alpha)_{i'}^i \delta_{j'}^j - \delta_i^{i'} (\tau_\alpha)_{j'}^j. \quad (132.20)$$

To decompose it into irreducible representations it suffices to separate from the matrix φ_i^j the non-zero trace, i.e. to rewrite it in the form

$$\varphi_i^j \equiv (\varphi_i^j - \frac{1}{2} \delta_i^j \varphi_k^k) + \frac{1}{2} \delta_i^j \varphi_k^k. \quad (132.21)$$

The last term is invariant with respect to transformations from the group SU(2), i.e. is a scalar. The trace of the matrix standing in parentheses is equal to zero. It contains three independent components transforming according to the three-dimensional irreducible representation which is called the vector representation. Symbolically the expansion (132.21) is written in the form

$$(1,0) \otimes (0,1) = (1,1) \oplus (0,0). \quad (132.22)$$

From (132.15) and (132.8) it follows that for the group SU(2) the quantities \hat{L}^2 are invariant operators:

$$[\hat{L}^2, \hat{L}_\alpha] = 0 \quad \text{where} \quad \hat{L}^2 = \hat{L}_1^2 + \hat{L}_2^2 + \hat{L}_3^2. \quad (132.23)$$

Making use of (132.16) and (132.7) it can be shown that any matrix $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ of the multiplet (p,q) is an eigenfunction of the operator \hat{L}^2 , where

$$\hat{L}^2 \varphi_{i_1 \dots i_p}^{j_1 \dots j_q} = (p+q)(p+q+2) \varphi_{i_1 \dots i_p}^{j_1 \dots j_q}. \quad (132.24)$$

The eigenvalues of the operator \hat{L}^2 depend only on the type of multiplet (more precisely, only on its dimension) and are its characteristics.

To classify the basis elements of a multiplet, which can be chosen to be combinations of the matrices $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ with only one non-zero element, use can be made of the eigenvalues of the diagonal operator \hat{L}_3 . Since the rank of the group SU(2) is equal to 1, this quantum number is sufficient. From (132.16) and (132.7) we conclude that if among p subscripts there are p_2 and among q superscripts q_2 indices equal to 2, then

$$\hat{L}_3 \varphi_{i_1 \dots i_p}^{j_1 \dots j_q} = [(p-q) - 2(p_2 - q_2)] \varphi_{i_1 \dots i_p}^{j_1 \dots j_q}. \quad (132.25)$$

The neighbouring eigenvalues of the generator \hat{L}_3 , which correspond to the basis elements of a given multiplet, differ by 2. From (132.25) it is seen that the minimum of these is equal to $-(p+q)$ and the maximum to $+(p+q)$, so that there are in all $p+q+1$ independent terms of the multiplet, and we again arrive at formula (132.17) for the dimension of the irreducible representation.

§133. Isomultiplets of elementary particles

We now suppose that the strong interaction of elementary particles is invariant with respect to the group $SU(2)$. This means, first of all, that the wave functions of hadrons transform according to certain of its irreducible representations, i.e. they are the products of the ordinary wave function, depending on spatial coordinates and on the spin projection, and the isomatrix:

$$\Psi = \psi(x, y, z, s_z) \phi_{i_1 \dots i_p}^{j_1 \dots j_q}, \quad (133.1)$$

In this case an individual hadron can conveniently be compared with a wave function in which the matrix $\phi_{i_1 \dots i_p}^{j_1 \dots j_q}$ is one of the basis elements. Thus all hadrons turn out to be distributed in isomultiplets, which are characterized by the eigenvalues of the generator L^2 . Individual hadrons within an isomultiplet are classified by the eigenvalues of the generator L_3 . In consequence of the assumed invariance of the strong interaction with respect to the isogroup, these two quantum numbers will be conserved in all reactions due to it. Furthermore, quantum-mechanical degeneracy must occur, i.e. the masses of the particles belonging to one isomultiplet will be strictly the same. When the electromagnetic interaction, which is not considered to be invariant with respect to the isogroup, is switched on, this degeneracy is removed and the isomultiplets split into individual particles with somewhat different masses. An analogue of this is, for example, the Zeeman effect (§74), in which the application of an external magnetic field violating the invariance with respect to the rotation group removes the earlier degeneracy of levels with respect to the angular momentum component J_z .

We note that for hadrons involved in the same isomultiplet all quantum numbers (spin σ , parity P , baryonic number B and strangeness S or hypercharge Y) differing from the eigenvalues of L_3 must be the same. This follows from the fact that the corresponding operators commute with the generators L_α (i.e. are invariant operators), and from the Shur lemma.

Instead of the generators L_α it is convenient to introduce the operators

$$\hat{T}_\alpha = \frac{1}{2} L_\alpha \quad (133.2)$$

which are called isospin component operators. Then as an invariant operator it is natural to take the operator of the square of the isospin

$$\hat{T}^2 = \hat{T}_1^2 + \hat{T}_2^2 + \hat{T}_3^2 = \frac{1}{4} L^2. \quad (133.3)$$

Formulae (132.24) and (132.25) now assume the form

$$\hat{T}^2 \varphi_{i_1 \dots i_p}^{j_1 \dots j_q} = \frac{1}{4}(p+q)(p+q+2) \varphi_{i_1 \dots i_p}^{j_1 \dots j_q} \equiv T(T+1) \varphi_{i_1 \dots i_p}^{j_1 \dots j_q}, \quad (133.4)$$

$$\hat{T}_3 \varphi_{i_1 \dots i_p}^{j_1 \dots j_q} = [\frac{1}{2}(p-q) - (p_2 - q_2)] \varphi_{i_1 \dots i_p}^{j_1 \dots j_q} \equiv T_3 \varphi_{i_1 \dots i_p}^{j_1 \dots j_q}. \quad (133.5)$$

The quantum number

$$T = \frac{1}{2}(p+q) \quad (133.6)$$

classifies the hadron multiplets and is called the isospin, and the quantity

$$T_3 = \frac{1}{2}(p-q) - (p_2 - q_2) \quad (133.7)$$

classifies the basis elements of the isomultiplet, i.e. the individual hadrons belonging to it. It can take on values from $-T$ up to $+T$, and is said to be the third component of isospin. Formula (132.17) for the dimension of the irreducible representation, i.e. for the number of different particles involved in the isomultiplet, takes the form

$$N(p, q) \equiv N(T) = 2T + 1. \quad (133.8)$$

For the wave functions which are the eigenfunctions of the operators \hat{T}^2 and \hat{T}_3 with eigenvalues T and T_3 we shall henceforth sometimes use Dirac's notation $|T, T_3\rangle$. Finally, in correspondence with the empirical Gell-Mann–Nishijima relation (§129) we shall assume by definition that the charge operator is

$$\hat{Q} = \hat{T}_3 + \frac{1}{2}(B+S)\hat{I} = \hat{T}_3 + \frac{1}{2}Y\hat{I} \quad (133.9)$$

where \hat{I} is the unit operator in the space of functions $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$.

The above relations point to a close relation of the group $SU(2)$ to the rotation group $O(3)$, the isospin T being a complete analogue of the angular momentum J . This relationship allows one to make use of the entire mathematical apparatus of angular momentum theory presented in §§51 and 52. In particular, the formalism of the Clebsch–Gordan coefficients (§52) is applicable for the actual decomposition of the direct product of representations into irreducible representations.

Let us now consider actual isomultiplets of hadrons. The proton and neutron have the same masses and are identical with respect to strong interactions. They form an isodoublet, i.e. their wave functions are of the form

$$\Psi_p = \psi_p \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \psi_p \\ 0 \end{pmatrix} \equiv \begin{pmatrix} p \\ 0 \end{pmatrix} \equiv N_1; \quad \Psi_n = \psi_n \begin{pmatrix} 0 \\ 1 \end{pmatrix} \equiv \begin{pmatrix} 0 \\ n \end{pmatrix} \equiv N_2. \quad (133.10)$$

The wave function of a nucleon is written as

$$N_i = \begin{pmatrix} p \\ n \end{pmatrix}. \quad (133.11)$$

Under the normalization condition

$$\sum_{\sigma_z = \pm \frac{1}{2}} \int dV (|p|^2 + |n|^2) = 1 \quad (133.12)$$

the two terms of this expression corresponding to the proton and neutron are interpreted as the probabilities of observing the nucleon in the proton state and in the neutron state. The spinor representation has dimension 2, so that the nucleon isodoublet is to be assigned isospin $\frac{1}{2}$. The generators of the representation are the matrices τ_α , so that the isospin components T_3 are the eigenvalues of the operator $\frac{1}{2}\tau_3$. To find them use can be made of formula (133.7), but in the given case they are easily determined directly:

$$\begin{aligned} \hat{T}_3 \begin{pmatrix} p \\ 0 \end{pmatrix} &= \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} p \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} p \\ 0 \end{pmatrix}, \\ \hat{T}_3 \begin{pmatrix} 0 \\ n \end{pmatrix} &= \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ n \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} 0 \\ n \end{pmatrix}, \end{aligned} \quad (133.13)$$

so that for the proton $T_3 = +\frac{1}{2}$, and for the neutron $T_3 = -\frac{1}{2}$. In correspondence with (133.8), taking into account that for the nucleon $B = 1$, $S = 0$, we obtain

$$\hat{Q}_N = \hat{T}_3 + \frac{1}{2}\hat{I} = \frac{1}{2}(\hat{\tau}_3 + \hat{I}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad (133.14)$$

so that

$$\begin{aligned} \hat{Q}_N \begin{pmatrix} p \\ 0 \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} p \\ 0 \end{pmatrix} = \begin{pmatrix} p \\ 0 \end{pmatrix}, \\ \hat{Q}_N \begin{pmatrix} 0 \\ n \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned} \quad (133.15)$$

Thus the charges of the proton and neutron are equal respectively to +1 and 0, as is to be expected.

The Ξ -hyperon, the Ξ_{1530} resonance, the K-meson and the K^* -meson resonance are also isodoublets. The relations written above are valid for them, with the only difference that for the Ξ -hyperon and its resonance the hypercharge is $Y = -1$ ($B=1, S=-2$), so that for these particles the charge operator is equal to

$$\hat{Q}_{\Xi} = \hat{T}_3 - \frac{1}{2}\hat{I} = \frac{1}{2}(\hat{\tau}_3 - \hat{I}) = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}. \quad (133.16)$$

Thus there are the following hadron isodoublets:

$$\begin{aligned} N_i &= \begin{pmatrix} p \\ n \end{pmatrix}; & \Xi_i &= \begin{pmatrix} \Xi^0 \\ \Xi^- \end{pmatrix}, & (\Xi_{1530})_i &= \begin{pmatrix} \Xi_{1530}^0 \\ \Xi_{1530}^- \end{pmatrix}, \\ K_i &= \begin{pmatrix} K^+ \\ K^0 \end{pmatrix}, & K_i^* &= \begin{pmatrix} K^{*+} \\ K^{*0} \end{pmatrix}. \end{aligned} \quad (133.17)$$

It is natural to assume that the wave functions of the antiproton and antineutron transform according to the representation conjugate to the nucleon representation, i.e.

$$\begin{aligned} \Psi_{\bar{p}} &= \psi_{\bar{p}}(1, 0) = (\psi_{\bar{p}}, 0) \equiv (\bar{p}, 0), \\ \Psi_{\bar{n}} &= \psi_{\bar{n}}(0, 1) = (0, \bar{n}). \end{aligned} \quad (133.18)$$

Hence the wave function of the antinucleon is written in the form

$$\bar{N}^i = (\bar{p}, \bar{n}). \quad (133.19)$$

The isospin of the antinucleon is also equal to $\frac{1}{2}$, but, because all the generators have now changed sign, $\hat{T}_3 = -\frac{1}{2}\hat{\tau}_3$ and hence for the antiproton $T_3 = -\frac{1}{2}$, and for the antineutron $T_3 = +\frac{1}{2}$ (it should be noted that, according to the rules of matrix multiplication by a wave function having the form of a row, the operators which are square matrices act from the right and not from the left). It is obvious that the antinucleon charge operator $\hat{Q}_{\bar{N}}$ is equal to $-\hat{Q}_{\bar{N}}$ so that the antiproton has $Q = -1$ and the antineutron $Q = 0$. All these statements may also be applied in an obvious way to the isodoublets of other antiparticles.

There exist three π -mesons: π^+ , π^0 and π^- , with about the same masses, identical with respect to the strong interaction. It is natural to assume that

they form an isotriplet, i.e. their wave functions transform according to the vector representation, forming a second-order matrix with trace zero:

$$\pi_i^j = \begin{pmatrix} \pi_1^1 & \pi_1^2 \\ \pi_2^1 & \pi_2^2 \end{pmatrix}. \quad (133.20)$$

The condition $\text{Tr } \pi_i^j = 0$ requires that

$$\pi_1^1 = -\pi_2^2. \quad (133.21)$$

To the isotriplet there corresponds the isospin $T = 1$. Making use of formula (133.7) we find that to the components π_1^1 , π_1^2 and π_2^2 there correspond respectively $T_3 = 0$, $T_3 = +1$ and $T_3 = -1$. Since for π -mesons $B = S = 0$, then

$$\hat{Q}_\pi = \hat{T}_3 \quad (133.22)$$

i.e. their charges are the same as the values of the component T_3 . Hence

$$\pi_1^1 = -\pi_2^2 \sim \pi^0, \quad \pi_1^2 \sim \pi^+, \quad \pi_2^2 \sim \pi^-. \quad (133.23)$$

By requiring that the isomatrices involved in the wave function of each π -meson be normalized to one, i.e. that they form the orthonormalized basis of the isotriplet, and taking into account (133.20) and (133.23), we finally obtain for the total wave function of the π -meson

$$\pi_i^j = \begin{pmatrix} \pi^0/\sqrt{2} & \pi^+ \\ \pi^- & -\pi^0/\sqrt{2} \end{pmatrix}. \quad (133.24)$$

Taking into account that the components π_i^j transform in terms of each other according to the law (132.18) and making use of the properties of unitary unimodular matrices, it is easily shown that the wave functions

$$\pi^+ = \pi_1^2, \quad \pi^0 = (\pi_1^1 - \pi_2^2)/\sqrt{2}, \quad \pi^- = \pi_2^1 \quad (133.25)$$

transform as the ordinary components of a three-dimensional vector. This accounts for the name of the representation (132.1). Hence the wave function of the π -meson can be written in the form

$$\pi = \begin{pmatrix} \pi^+ \\ \pi^0 \\ \pi^- \end{pmatrix}. \quad (133.26)$$

In such a formalism the generators of the vector representation and the isospin component operators \hat{T}_3 will be Hermitian 3×3 matrices with trace zero. Making use of (51.17) we can write immediately

$$\hat{T}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}; \quad \hat{T}_2 = \frac{1}{2} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}; \quad \hat{T}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (133.27)$$

By direct check we see indeed that

$$\hat{T}_3 \begin{pmatrix} \pi^+ \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \pi^+ \\ 0 \\ 0 \end{pmatrix}; \quad \hat{T}_3 \begin{pmatrix} 0 \\ \pi^0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \pi^0 \\ 0 \end{pmatrix}; \quad \hat{T}_3 \begin{pmatrix} 0 \\ 0 \\ \pi^- \end{pmatrix} = - \begin{pmatrix} 0 \\ 0 \\ \pi^- \end{pmatrix}. \quad (133.28)$$

The results obtained apply automatically also to other hadron isotriplets: Σ , ρ , Σ_{1385} . Since in these cases $Y=0$, the charge operator also has the form (133.22).

Since the representation conjugate to (1,1) coincides with itself, and their generators are also the same, the structure of antiparticle isotriplets is identical with that of the matrix (133.24). For example,

$$\bar{\Sigma}_i' = \begin{pmatrix} \bar{\Sigma}^0/\sqrt{2} & \bar{\Sigma}^- \\ \bar{\Sigma}^+ & -\bar{\Sigma}^0/\sqrt{2} \end{pmatrix} \quad (133.29)$$

(we note that the charge of $\bar{\Sigma}^-$ is positive, and that of $\bar{\Sigma}^+$ is negative). For mesons there are no quantum numbers, except of the component T_3 , which would distinguish between particles and antiparticles (for the Σ -hyperon such quantum numbers are the baryonic number and strangeness). Hence the antiparticle with respect to π^+ is π^- and vice versa, while in the case of the π^0 -meson the particle is identical with the antiparticle. This accounts for the fact that for π -mesons the particles and antiparticles are contained in the same isomultiplet, whereas, say, Σ and $\bar{\Sigma}$ form two different isomultiplets. An analogous situation also occurs in the case of ρ -mesons

$$\rho_i' = \begin{pmatrix} \rho^0/\sqrt{2} & \rho^+ \\ \rho^- & -\rho^0/\sqrt{2} \end{pmatrix} \quad \text{where} \quad \bar{\rho}^+ = \rho^-; \quad \bar{\rho}^- = \rho^+; \quad \bar{\rho}^0 = \rho^0. \quad (133.30)$$

The four nucleon resonances Δ^{++} , Δ^+ , Δ^0 and Δ^- form an isoquartet i.e. their wave functions transform according to the irreducible representation of dimension 4, forming the matrix Δ_{ijk} symmetric in any pair of indices. To the isoquartet there corresponds the isospin $T = \frac{3}{2}$. From (133.7) it follows that for the components Δ_{111} , $\Delta_{112} = \Delta_{121} = \Delta_{211}$, $\Delta_{212} = \Delta_{221} = \Delta_{122}$ and Δ_{222} the isospin component T_3 is equal respectively to $+\frac{3}{2}$, $+\frac{1}{2}$, $-\frac{1}{2}$ and $-\frac{3}{2}$. Since for the Δ -resonances $Y=1$ ($B=1, S=0$), then

$$\hat{Q}_\Delta = \hat{T}_3 + \frac{1}{2}\hat{I} \quad (133.31)$$

so that

$$\begin{aligned} \Delta_{111} &\sim \Delta^{++}, & \Delta_{112} = \Delta_{121} = \Delta_{211} &\sim \Delta^+, \\ \Delta_{122} = \Delta_{212} = \Delta_{221} &\sim \Delta^0, & \Delta_{222} &\sim \Delta^-. \end{aligned} \quad (133.32)$$

From the normalization condition it follows that in the first and last relations the factor of proportionality is equal to 1, while in the second and third cases it is equal to $1/\sqrt{3}$. The isoquartet $\bar{\Delta}$ is filled in an obvious way.

There are hadrons, for example Λ - and Ω -hyperons and η -, ω - and φ -mesons, which are isosinglets, i.e. their wave functions transform according to the isoscalar representation. Since its generators are equal to zero, $T = T_3 = 0$ for an isosinglet and $Q = \frac{1}{2}Y$. The antiparticles $\bar{\eta}$, $\bar{\omega}$, $\bar{\varphi}$ are the same as the corresponding particles.

§ 134. The wave functions of a system of nucleons and π -mesons

Let us now consider the three simplest composite systems: NN , $N\bar{N}$ and πN , which are of considerable interest.

1. We decompose the wave function of the system nucleon-antinucleon

$$\varphi_i^j = \bar{N}^j N_i \quad (134.1)$$

into the irreducible parts (§ 132):

$$\varphi_i^j = (\bar{N}^j N_i - \frac{1}{2} \delta_i^j \bar{N}^k N_k) + \frac{1}{2} \delta_i^j \bar{N}^k N_k \equiv \chi_i^j + \delta_i^j \chi. \quad (134.2)$$

The last term is an isoscalar, while the expression standing in parentheses is an isovector. Recalling the analysis of the matrix (133.20) we obtain

$$\begin{aligned} |1, +1\rangle &= a\chi_1^2 = a\bar{n}p, & |1, 0\rangle &= b\chi_1^1 = b\frac{1}{2}(\bar{p}p - \bar{n}n), \\ |1, -1\rangle &= c\chi_2^1 = c\bar{p}n, & |0, 0\rangle &= d\chi = d\frac{1}{2}(\bar{p}p + \bar{n}n) \end{aligned} \quad (134.3)$$

where a , b , c and d are normalization coefficients. Assuming that the spatial-spin parts of all the wave functions \bar{N}^j and N_i are normalized to unity, from the condition of orthonormalization of the basis states (134.3) we have $a^2 = c^2 = 1$, $b^2 = d^2 = 2$. Thus, finally,

$$\begin{aligned}
 |1, +1\rangle &= \bar{n}p, & |1, 0\rangle &= \frac{1}{\sqrt{2}}(\bar{p}p - \bar{n}n), \\
 |1, -1\rangle &= \bar{p}n, & |0, 0\rangle &= \frac{1}{\sqrt{2}}(\bar{p}p + \bar{n}n).
 \end{aligned}
 \tag{134.4}$$

If the system $N\bar{N}$ is in the 1S_0 -state (spins antiparallel), then its total spin is equal to 0, and the parity is odd (the relative parities of the particle and antiparticle are opposite). Thus one can construct from the nucleon and antinucleon a pseudoscalar isotriplet and a pseudoscalar isosinglet, i.e. a π -meson and an η -meson. This result underlies the composite model of the π -meson, proposed in 1949 by Fermi and Yang. The isotriplet state 3S_1 of the $N\bar{N}$ pair can be compared to the ρ -meson.

2. Considering the wave function of a system of two nucleons

$$\varphi_{ij} = N'_i N''_j \tag{134.5}$$

we can separate it into parts symmetric and antisymmetric with respect to the indices:

$$\varphi_{ij} = N'_i N''_j = \frac{1}{2}(N'_i N''_j + N'_j N''_i) + \frac{1}{2}(N'_i N''_j - N'_j N''_i) \equiv \chi_{[ij]} + \chi_{\{ij\}}. \tag{134.6}$$

The second term contains one non-zero independent component ($i=1, j=2$), i.e. an isoscalar, while the first term is an isovector. Making use of formula (133.7), we obtain

$$\begin{aligned}
 |1, +1\rangle &= a\chi_{11} = ap'p'', & |1, 0\rangle &= b\chi_{12} = b\frac{1}{2}(p'n'' + n'p''), \\
 |1, -1\rangle &= c\chi_{22} = cn'n'', & |0, 0\rangle &= d\chi_{\{12\}} = d\frac{1}{2}(p'n'' - n'p'').
 \end{aligned}
 \tag{134.7}$$

The normalization coefficients are defined as before. We have

$$\begin{aligned}
 |1, +1\rangle &= p'p'', & |1, 0\rangle &= \frac{1}{\sqrt{2}}(p'n'' + n'p''), \\
 |1, -1\rangle &= n'n'', & |0, 0\rangle &= \frac{1}{\sqrt{2}}(p'n'' - n'p'')
 \end{aligned}
 \tag{134.8}$$

(see relations (66.4) and (66.5)). The first three functions, corresponding to isospin 1, are symmetric, whereas the last function, corresponding to the isoscalar system pn , is antisymmetric in the isovariables.

In the given formalism the proton and neutron are considered as two states of one particle, the nucleon. Hence the total wave function of a system of two nucleons considered as identical particles must possess definite symmetry properties with respect to their exchange. Since the type of symmetry does not depend on what pair of particles is exchanged, we exchange

two protons. They obey the Pauli statistics, i.e. when their coordinates and spins are exchanged the wave function changes sign. On the other hand, from (134.8) it is seen that when the isovariables of two protons are exchanged the wave function does not change. Thus nucleons obey the generalized Pauli principle according to which the total wave function of a system of nucleons is antisymmetric with respect to the exchange of any of their pairs. Hence it follows, in particular, that the wave function of a system NN with $T=1$ describes a state whose angular momentum differs from that of the isoscalar ($T=0$) state.

3. From the wave function of a system of a π -meson and a nucleon

$$\varphi_{ij}^k = N_i \pi_j^k \quad (134.9)$$

we separate the parts symmetric and antisymmetric with respect to subscripts:

$$\varphi_{ij}^k = \frac{1}{2}(N_i \pi_j^k + N_j \pi_i^k) + \frac{1}{2}(N_i \pi_j^k - N_j \pi_i^k) \equiv \varphi_{[ij]}^k + \varphi_{\{ij\}}^k. \quad (134.10)$$

The second term contains two non-zero independent components $\varphi_{[12]}$, i.e. it is an isospinor ($T=\frac{1}{2}$). But the first matrix is still reducible, since its traces are not equal to zero. Separating them, we get

$$\begin{aligned} \varphi_{[ij]}^k &= \frac{1}{2}(N_i \pi_j^k + N_j \pi_i^k - \frac{1}{3}\delta_i^k N_m \pi_j^m - \frac{1}{3}\delta_j^k N_m \pi_i^m) + \\ &+ \frac{1}{6}(\delta_i^k N_m \pi_j^m + \delta_j^k N_m \pi_i^m) \equiv \chi_{ij}^k + (\delta_i^k \chi_j + \delta_j^k \chi_i). \end{aligned} \quad (134.11)$$

The second term represents an isospinor. It is easily seen that its components are the same as the two components $\varphi_{\{12\}}^k$. From (133.6) we conclude that the first term corresponds to isospin $T=\frac{3}{2}$, i.e. that it contains four independent components. Making use of (133.7), we have

$$\begin{aligned} |\frac{3}{2}, +\frac{3}{2}\rangle &\sim \chi_{11}^2, & |\frac{3}{2}, +\frac{1}{2}\rangle &\sim \chi_{11}^1, & |\frac{3}{2}, -\frac{1}{2}\rangle &\sim \chi_{22}^2, \\ |\frac{3}{2}, -\frac{3}{2}\rangle &\sim \chi_{22}^1, & |\frac{1}{2}, +\frac{1}{2}\rangle &\sim \chi_1, & |\frac{1}{2}, -\frac{1}{2}\rangle &\sim \chi_2. \end{aligned} \quad (134.12)$$

Writing the components explicitly and defining the factors of proportionality from the normalization conditions, we finally obtain

$$\begin{aligned} |\frac{3}{2}, +\frac{3}{2}\rangle &= p\pi^+, & |\frac{3}{2}, +\frac{1}{2}\rangle &= \sqrt{\frac{2}{3}}p\pi^0 - \sqrt{\frac{1}{3}}n\pi^+, & |\frac{3}{2}, -\frac{1}{2}\rangle &= \sqrt{\frac{1}{3}}p\pi^- + \sqrt{\frac{2}{3}}n\pi^0, \\ |\frac{3}{2}, -\frac{3}{2}\rangle &= n\pi^-, & |\frac{1}{2}, +\frac{1}{2}\rangle &= \sqrt{\frac{1}{3}}p\pi^0 + \sqrt{\frac{2}{3}}n\pi^+, & |\frac{1}{2}, -\frac{1}{2}\rangle &= \sqrt{\frac{2}{3}}p\pi^- - \sqrt{\frac{1}{3}}n\pi^0. \end{aligned} \quad (134.13)$$

One can also arrive at this result by means of the formalism developed in §§51 and 52. According to the rules of addition of angular momenta the

isospin of a system consisting of a π -meson ($T=1$) and a nucleon ($T=\frac{1}{2}$) can take on the values $\frac{3}{2}$ and $\frac{1}{2}$. To construct the wave functions of the system corresponding to definite values of T and T_3 use can be made of the general formula (52.3), which in our notation is written as

$$|T, T_3\rangle = \sum_{t_3=\pm\frac{1}{2}} C_{T_3-t_3, t_3}^T |1, T_3-t_3\rangle |\frac{1}{2}, t_3\rangle. \quad (134.14)$$

Taking the Clebsch–Gordan coefficients from the table 1, §52, we have

$$\begin{aligned} |\frac{3}{2}, +\frac{3}{2}\rangle &= C_{1, \frac{1}{2}}^{\frac{3}{2}} |1, +1\rangle |\frac{1}{2}, +\frac{1}{2}\rangle = p\pi^+, \\ |\frac{3}{2}, +\frac{1}{2}\rangle &= C_{1, -\frac{1}{2}}^{\frac{3}{2}} |1, +1\rangle |\frac{1}{2}, -\frac{1}{2}\rangle + C_{0, \frac{1}{2}}^{\frac{3}{2}} |1, 0\rangle |\frac{1}{2}, +\frac{1}{2}\rangle = \sqrt{\frac{2}{3}}p\pi^0 - \sqrt{\frac{1}{3}}n\pi^+, \\ |\frac{3}{2}, -\frac{1}{2}\rangle &= C_{0, -\frac{1}{2}}^{\frac{3}{2}} |1, 0\rangle |\frac{1}{2}, -\frac{1}{2}\rangle + C_{-1, \frac{1}{2}}^{\frac{3}{2}} |1, -1\rangle |\frac{1}{2}, +\frac{1}{2}\rangle = \sqrt{\frac{1}{3}}p\pi^- + \sqrt{\frac{2}{3}}n\pi^0, \\ |\frac{3}{2}, -\frac{3}{2}\rangle &= C_{-1, -\frac{1}{2}}^{\frac{3}{2}} |1, -1\rangle |\frac{1}{2}, -\frac{1}{2}\rangle = n\pi^-, \\ |\frac{1}{2}, +\frac{1}{2}\rangle &= C_{1, -\frac{1}{2}}^{\frac{1}{2}} |1, +1\rangle |\frac{1}{2}, -\frac{1}{2}\rangle + C_{0, \frac{1}{2}}^{\frac{1}{2}} |1, 0\rangle |\frac{1}{2}, +\frac{1}{2}\rangle = \sqrt{\frac{1}{3}}p\pi^0 + \sqrt{\frac{2}{3}}n\pi^+, \\ |\frac{1}{2}, -\frac{1}{2}\rangle &= C_{0, -\frac{1}{2}}^{\frac{1}{2}} |1, 0\rangle |\frac{1}{2}, -\frac{1}{2}\rangle + C_{-1, \frac{1}{2}}^{\frac{1}{2}} |1, -1\rangle |\frac{1}{2}, +\frac{1}{2}\rangle = \sqrt{\frac{2}{3}}p\pi^- - \sqrt{\frac{1}{3}}n\pi^0. \end{aligned}$$

Thus we again arrive at relations (134.13).

One can easily reverse them and express the wave function of a system consisting of a π -meson and a nucleon in terms of the functions $|T, T_3\rangle$:

$$\begin{aligned} p\pi^+ &= |\frac{3}{2}, +\frac{3}{2}\rangle, \\ p\pi^- &= \sqrt{\frac{1}{3}}|\frac{3}{2}, -\frac{1}{2}\rangle + \sqrt{\frac{2}{3}}|\frac{1}{2}, -\frac{1}{2}\rangle, \\ p\pi^0 &= \sqrt{\frac{2}{3}}|\frac{3}{2}, +\frac{1}{2}\rangle + \sqrt{\frac{1}{3}}|\frac{1}{2}, +\frac{1}{2}\rangle, \\ n\pi^+ &= -\sqrt{\frac{1}{3}}|\frac{3}{2}, +\frac{1}{2}\rangle + \sqrt{\frac{2}{3}}|\frac{1}{2}, +\frac{1}{2}\rangle, \\ n\pi^- &= |\frac{3}{2}, -\frac{3}{2}\rangle, \\ n\pi^0 &= \sqrt{\frac{2}{3}}|\frac{3}{2}, -\frac{1}{2}\rangle - \sqrt{\frac{1}{3}}|\frac{1}{2}, -\frac{1}{2}\rangle. \end{aligned} \quad (134.15)$$

§135. Isotopically invariant interaction

In §67 we pointed out that nuclear forces, i.e. the forces acting between two nucleons of the atomic nucleus, possess the property of charge independence. This property is most simply described within the framework of the isospin formalism. The hypothesis of charge independence of nuclear

forces says that the interaction of two protons, two neutrons and of a proton with a neutron in the same spatial-spin states is identical. From the generalized Pauli principle it follows that if the two nucleons are in the isotriplet state (symmetric with respect to isospin variables), then their coordinate wave function is antisymmetric. For the proton and neutron in the isosinglet state, the spin-coordinate part of the wave function is symmetric. This means that the interaction of two nucleons in states $|1,+1\rangle$, $|1,0\rangle$ and $|1,-1\rangle$, other things being equal, will be the same, whereas the interaction in state $|0,0\rangle$ is, in general, essentially different.

As an example let us consider the deuteron, which consists of a proton and a neutron. The coordinate wave function of its ground state ($l=0$) is symmetric, and the spin wave function is also symmetric; the total angular momentum of the deuteron is equal to 1, i.e. it is in the triplet spin state 3S_1 with a small admixture of state 3D_1 (see §76). Hence the isospin part of the wave function must be antisymmetric and form an isosinglet. It is known from experiment that the binding energy of the deuteron is equal to 2.23 MeV. On the other hand, if the deuteron were in the state 1S_0 , then its isospin wave function would be symmetric, i.e. it would be an isovector. In this case there are no bound states.

Thus, under the assumption of charge independence, nuclear forces are determined by the total isospin T , rather than by its component T_3 . Consequently, in the phenomenological theory of nuclear forces, in which the interaction between nucleons is described by a certain potential, the Hamiltonian may contain only the invariant operator \hat{T}^2 , the square of the total isospin of the system of nucleons, which can be written in the form

$$\hat{T}^2 = (\hat{\mathbf{T}}' + \hat{\mathbf{T}}'')^2 = \frac{1}{4}(\boldsymbol{\tau}' + \boldsymbol{\tau}'')^2 = \frac{1}{4}\boldsymbol{\tau}'^2 + \frac{1}{4}\boldsymbol{\tau}''^2 + \frac{1}{2}(\boldsymbol{\tau}' \cdot \boldsymbol{\tau}''). \quad (135.1)$$

Here $\boldsymbol{\tau}' = \{\tau'_1, \tau'_2, \tau'_3\}$ and $\boldsymbol{\tau}'' = \{\tau''_1, \tau''_2, \tau''_3\}$ are Pauli matrices acting on the isospin indices of the first and second nucleons respectively. Since the operators $\boldsymbol{\tau}'^2$ and $\boldsymbol{\tau}''^2$ are multiples of the unit operator, the dependence of the interaction Hamiltonian on the isospin variables is defined only by the scalar product $(\boldsymbol{\tau}' \cdot \boldsymbol{\tau}'')$:

$$\hat{H}_{\text{int}} = U_1 + U_2(\boldsymbol{\tau}' \cdot \boldsymbol{\tau}''). \quad (135.2)$$

Here U_1 and U_2 depend only on the coordinates and on the ordinary spin; they are written down in §76.

It is easily verified that the Hamiltonian (135.2) commutes with the isospin component operators \hat{T}_α , and thus also with the operator \hat{T}^2 :

$$[\hat{H}_{\text{int}}, \hat{T}_\alpha] = [\hat{H}_{\text{int}}, \hat{T}^2] = 0. \quad (135.3)$$

Consequently, if the Coulomb interaction of protons and the small difference in the masses of the proton and neutron are disregarded, then in the system of interacting nucleons there holds not only the law of conservation of the isospin component, expressing the trivial fact of charge conservation, but also the law of conservation of the total isospin, which may serve as a formulation of charge independence.

Let us now turn to the interaction of nucleons with π -mesons. As we know from §67, it is just this interaction which is responsible for the existence of nuclear forces, i.e. for the interaction of nucleons, which we have considered above purely phenomenologically. A logical discussion of the corresponding problems is possible only within the framework of the quantum field theory. Therefore we shall confine ourselves to some comments. Analogous to what we did in §122 in describing the weak interaction, we shall consider the nucleon function N_i and the π -meson function π_j^k as operators in the space of occupation numbers. The operator N_i corresponds to the destruction of a nucleon and to the creation of an antinucleon, whereas the operator $(N^\dagger)_i$ creates a nucleon and destroys an antinucleon. The operator π_j^k creates and destroys π -mesons.

The basic processes of the interaction considered are those of virtual production and absorption of π -mesons by nucleons (see §67). Hence the interaction Hamiltonian density must have the general structure $N^\dagger N \pi$. Since the π -meson operator is a pseudoscalar, relativistic invariance and parity conservation require that it be multiplied by the pseudoscalar combination of N^\dagger and N . Recalling the results of §121, we arrive at the expression $\bar{N} \gamma_5 N \pi$. In consequence of the charge independence of the πN interaction, the Hamiltonian density must be an isoscalar. One can form from the matrices \bar{N}^i , N_j and π_k^l only one combination of this type: $\bar{N}^i N_j \pi_i^j$. Thus, finally,

$$\hat{H}_{\text{int}} = \sqrt{2} g \bar{N}^i \gamma_5 N_j \pi_i^j \quad (135.4)$$

where g is the strong πN interaction constant (an analogue of the electric charge). The factor $\sqrt{2}$ is introduced for historical reasons.

Making use of the explicit form of the nucleon matrices and the π -meson matrices \bar{N}^i , N_j and π_k^l (see (133.17), (133.18) and (133.24)), from (135.4) we obtain

$$\hat{H}_{\text{int}} = g [\sqrt{2} \bar{p} \gamma_5 n \pi^+ + \sqrt{2} \bar{n} \gamma_5 p \pi^- + (\bar{p} \gamma_5 p - \bar{n} \gamma_5 n) \pi^0]. \quad (135.5)$$

Consequently, on the assumption of charge independence the following relation holds between the constants of the πN interaction:

$$g_{pn\pi^+} : g_{pn\pi^-} : g_{pp\pi^0} : g_{nn\pi^0} = 1 : 1 : \frac{1}{\sqrt{2}} : \left(-\frac{1}{\sqrt{2}}\right). \quad (135.6)$$

Taking into account the explicit form of the Pauli matrices and passing to the π -meson vector function π (133.26), it is easily verified that the Hamiltonian density (135.4) can be rewritten as

$$\hat{H}_{\text{int}} = g\bar{N}\gamma_5\tau N\pi \quad (135.7)$$

where it is assumed that the isovectors τ and π are multiplied in the scalar way. In older studies only this form of notation was used, which accounts for the appearance of the factor $\sqrt{2}$ in expression (135.4). The interaction of other baryons with mesons is described in an analogous way. For example, for the system π , Σ , $\bar{\Sigma}$ it is easily found that

$$\hat{H}_{\text{int}} = \sqrt{2}g'\bar{\Sigma}_i^j\gamma_5\Sigma_j^k\pi_k^i. \quad (135.8)$$

In conclusion we stress once more that the electromagnetic interaction of nucleons violates the invariance with respect to transformations of the isogroup SU(2), and the results formulated above are no longer valid. For this interaction the Hamiltonian density can be written on the basis of the same considerations as those which were used in obtaining (135.4). Taking into account that the operator of creation and destruction of photons is the vector potential A_μ , we shall have

$$\hat{H}_{\text{int}} = e\bar{N}\gamma_\mu\hat{Q}NA_\mu \quad (135.9)$$

where \hat{Q} is the nucleon charge operator given by formula (133.14). Hence

$$\hat{H}_{\text{int}} = \frac{1}{2}e\bar{N}\gamma_\mu(I+\tau_3)NA_\mu \quad (135.10)$$

i.e. the Hamiltonian density contains an isovector part (the term with τ_3) in addition to the isoscalar part. The presence of the former violates the conservation of the total isospin T , although its component T_3 is conserved. But the intensity of the electromagnetic interaction is much smaller than that of the strong interaction, so that electromagnetic corrections can frequently be neglected, being considered at most as a small perturbation.

Some quantitative consequences of the hypothesis of charge independence of the strong interaction, which is equivalent to total isospin conservation, are presented in the next section.

§136. The scattering of nucleons and π -mesons

We apply the isospin formalism to the analysis of the processes of scattering of nucleons by nucleons and of π -mesons by nucleons. Generalization to the case of other hadrons presents no difficulty.

Let several reactions

$$a_i + b_i \rightarrow c_i + d_i \quad (136.1)$$

be considered, all particles of the type a , b , c and d belonging to one and the same isomultiplet. For the wave functions of the initial and final states we shall make use of the Dirac notation $|a_i b_i\rangle$ and $|c_i d_i\rangle$. The scattering amplitude $f^{(i)}$ is proportional to the matrix element

$$M^{(i)} = \langle c_i d_i | a_i b_i \rangle \quad (136.2)$$

the square of the modulus of which defines the differential and, after integrating over angles, total cross sections for the process.

We first assume that the state of the particles before scattering has definite values of the isospin T and of its component T_3 , i.e. that its wave function is $|T, T_3\rangle$. If the part of the matrix element corresponding to Coulomb scattering is separated, then from charge independence it follows that in the reaction the isospin does not change:

$$\langle T', T_3 | T, T_3 \rangle = 0 \quad \text{for} \quad T' \neq T, \quad (136.3)$$

i.e. the wave function of the final state is also a function of the type $|T, T_3\rangle$. Furthermore, the matrix element corresponding to the scattering due to the strong interaction cannot depend on the values of the component T_3 , but is defined by the isospin T (and by other quantum numbers), by momenta, by spins and so on. We denote

$$\langle T, T_3 | T, T_3 \rangle \equiv M^{(T)}. \quad (136.4)$$

The validity of the isospin formalism as applied to the class of problems considered lies in the fact that the matrix element of any real process of the set (136.1) can be expressed in terms of a small number of (in most cases two) matrix elements $M^{(T)}$ corresponding to the scattering in a definite isospin state. For this it suffices to expand the wave functions $|a_i b_i\rangle$ and $|c_i d_i\rangle$ in terms of the wave functions $|T, T_3\rangle$, substitute these expansions into (136.2) and make use of formulae (136.3)–(136.4). Thus one can establish a number of relations between the cross sections for different processes corresponding to the same initial and the same final spatial-spin states of the particles involved in the scattering.

1. As the first example let us consider the scattering of protons by protons and of neutrons by protons. First of all, from (134.8) we express the wave functions of the initial and final states, i.e. of the systems pp or np, in terms of the basis functions of the isotriplet and isosinglet:

$$|p'p''\rangle = |1,+1\rangle, \quad |n'p''\rangle = \frac{1}{\sqrt{2}}(|1,0\rangle - |0,0\rangle), \quad |p'n''\rangle = \frac{1}{\sqrt{2}}(|1,0\rangle + |0,0\rangle). \quad (136.5)$$

Then for the process of scattering $p' + p'' \rightarrow p' + p''$ we find that

$$M^{(pp)} = \langle p'p'' | p'p'' \rangle = \langle 1,+1 | 1,+1 \rangle = M^{(1)}. \quad (136.6)$$

In neutron-proton scattering the following two processes are possible:
ordinary elastic scattering

$$n' + p'' \rightarrow n' + p''$$

and charge-exchange scattering

$$n' + p'' \rightarrow p' + n''.$$

For the first of these

$$M^{\text{elas}} = \langle n'p'' | n'p'' \rangle = \frac{1}{2} [(\langle 1,0 | - \langle 0,0 |)(|1,0\rangle - |0,0\rangle)] = \frac{1}{2} [M^{(1)} + M^{(0)}] \quad (136.7)$$

and for the second

$$M^{\text{ch.ex}} = \langle p'n'' | n'p'' \rangle = \frac{1}{2} [(\langle 1,0 | + \langle 0,0 |)(|1,0\rangle - |0,0\rangle)] = \frac{1}{2} [M^{(1)} - M^{(0)}]. \quad (136.8)$$

The elastic scattering cross sections are proportional to the squares of the moduli of the matrix elements, hence

$$\begin{aligned} \frac{d\sigma^{(pp)}}{d\Omega} &\sim |M^{(1)}|^2, & \frac{d\sigma^{\text{elas}}}{d\Omega} &\sim \frac{1}{4} |M^{(1)} + M^{(0)}|^2, \\ \frac{d\sigma^{\text{ch.ex}}}{d\Omega} &\sim \frac{1}{4} |M^{(1)} - M^{(0)}|^2. \end{aligned} \quad (136.9)$$

Summing the last two expressions, we obtain the total neutron-proton scattering cross section, which is determined experimentally by the total number of protons and neutrons scattered at a given angle. Finally

$$\frac{d\sigma^{(pp)}}{d\Omega} \sim |M^{(1)}|^2, \quad \frac{d\sigma^{(np)}}{d\Omega} \sim \frac{1}{2} |M^{(1)}|^2 + \frac{1}{2} |M^{(0)}|^2. \quad (136.10)$$

Since the angular dependence of the matrix elements $M^{(1)}$ and $M^{(0)}$ can be essentially different, then, in spite of charge independence, the behaviour of the proton-proton and neutron-proton scattering cross sections as func-

tions of the angular variable can be different. Experiment shows that this is indeed so. In the energy range of 300–500 MeV in the centre-of-mass system the first cross section is almost independent of the scattering angle, whereas the second has a minimum at $\theta = \frac{1}{2}\pi$, increasing sharply in the backward direction and to a lesser degree in the forward direction.

2. The example of the reaction of π -meson production with formation of a deuteron in nucleon–nucleon collisions is somewhat more interesting:

$$p + p \rightarrow d + \pi^+,$$

$$n + p \rightarrow d + \pi^0.$$

Since the deuteron isospin is equal to zero (see the beginning of § 135), then

$$|\pi^+ d\rangle = |1, +1\rangle, \quad |\pi^0 d\rangle = |1, 0\rangle. \quad (136.11)$$

Therefore

$$M^{(pp)} = \langle \pi^+ d | pp \rangle = \langle 1, +1 | 1, +1 \rangle = M^{(1)},$$

$$M^{(np)} = \langle \pi^0 d | np \rangle = \frac{1}{\sqrt{2}} [\langle 1, 0 | (|1, 0\rangle - |0, 0\rangle)] = \frac{1}{\sqrt{2}} M^{(1)}. \quad (136.12)$$

From this follows the relation between the cross sections:

$$\frac{d\sigma^{(pp)}/d\Omega}{d\sigma^{(np)}/d\Omega} = 2 \quad (136.13)$$

which has been confirmed experimentally.

3. The scattering of charged π -mesons by protons is an even more interesting case:

$$\pi^+ + p \rightarrow \pi^+ + p,$$

$$\pi^- + p \rightarrow \pi^- + p,$$

$$\pi^- + p \rightarrow \pi^0 + n.$$

Denoting the matrix elements and cross sections referring to these processes by the symbol of the π -meson in the final state and making use of formulae (134.15), we obtain

$$\begin{aligned}
 M^{(+)} &= \langle \pi^+ p | \pi^+ p \rangle = \langle \frac{3}{2}, +\frac{3}{2} | \frac{3}{2}, +\frac{3}{2} \rangle = M^{(\frac{3}{2})}, \\
 M^{(-)} &= \langle \pi^- p | \pi^- p \rangle = [(\sqrt{\frac{1}{3}} \langle \frac{3}{2}, -\frac{1}{2} | + \sqrt{\frac{2}{3}} \langle \frac{1}{2}, -\frac{1}{2} |) \times \\
 &\quad \times (\sqrt{\frac{1}{3}} \langle \frac{3}{2}, -\frac{1}{2} | + \sqrt{\frac{2}{3}} \langle \frac{1}{2}, -\frac{1}{2} |)] = \frac{1}{3} M^{(\frac{3}{2})} + \frac{2}{3} M^{(\frac{1}{2})}, \\
 M^{(0)} &= \langle \pi^0 n | \pi^- p \rangle = [(\sqrt{\frac{2}{3}} \langle \frac{3}{2}, -\frac{1}{2} | - \sqrt{\frac{1}{3}} \langle \frac{1}{2}, -\frac{1}{2} |) \times \\
 &\quad \times (\sqrt{\frac{1}{3}} \langle \frac{3}{2}, -\frac{1}{2} | + \sqrt{\frac{2}{3}} \langle \frac{1}{2}, -\frac{1}{2} |)] = \frac{\sqrt{2}}{3} M^{(\frac{3}{2})} - \frac{\sqrt{2}}{3} M^{(\frac{1}{2})},
 \end{aligned}$$

hence

$$f^{(+)} = f^{(\frac{3}{2})}, \quad f^{(-)} = \frac{1}{3} f^{(\frac{3}{2})} + \frac{2}{3} f^{(\frac{1}{2})}, \quad f^{(0)} = \frac{\sqrt{2}}{3} f^{(\frac{3}{2})} - \frac{\sqrt{2}}{3} f^{(\frac{1}{2})}. \quad (136.14)$$

Eliminating the amplitudes $f^{(\frac{3}{2})}$ and $f^{(\frac{1}{2})}$, we have the relation

$$f^{(+)} - f^{(-)} = \sqrt{2} f^{(0)} \quad (136.15)$$

from which the so-called triangle relations follow:

$$|\sqrt{\sigma^{(+)}} - \sqrt{\sigma^{(-)}}| \leq \sqrt{2\sigma^{(0)}} \leq \sqrt{\sigma^{(+)}} + \sqrt{\sigma^{(-)}}. \quad (136.16)$$

Under some additional assumptions regarding the properties of the amplitudes, more interesting relations can be obtained between the cross sections. Thus in the case $f^{(\frac{1}{2})} \approx 0$

$$\sigma^{(+)} : \sigma^{(-)} : \sigma^{(0)} = 9 : 1 : 2. \quad (136.17)$$

If it is assumed that $f^{(\frac{3}{2})} \approx f^{(\frac{1}{2})}$, then

$$\sigma^{(+)} : \sigma^{(-)} : \sigma^{(0)} = 1 : 1 : 0. \quad (136.18)$$

Finally, if $f^{(\frac{3}{2})} \approx 0$, then

$$\sigma^{(+)} : \sigma^{(-)} : \sigma^{(0)} = 0 : 2 : 1. \quad (136.19)$$

Experiment shows that for an energy of 120 MeV of the incident π -meson the total cross sections are in the ratio $93:11:22 \approx 9:1:2$, i.e. in this energy range the scattering in the state with isospin $T = \frac{3}{2}$ is dominating. For energies above 200 MeV the amplitude $f^{(\frac{1}{2})}$ also begins to give a considerable contribution.

There is also another, simpler method of obtaining the relations between the cross sections, which does not require knowledge of the Clebsch-Gordan coefficients and is especially useful in those cases where their calculation is for any reason difficult. It is called the method of invariant amplitudes, and will be demonstrated by the example of the scattering of charged π -mesons by nucleons.

Under the assumption of charge independence the isospin T does not

change in scattering. This means that the total amplitude of the scattering of one isomultiplet by another must be an isoscalar. In our case it is constructed of the wave functions N_i and π_i^j of the initial state and the wave functions \bar{N}^i and $\bar{\pi}_i^j$ of the final state which transform according to the conjugate representations (we denote by a bar the functions describing the final state; at the same time the bar is the symbol of the conjugate representation). The matrices of these wave functions are of the form

$$\begin{aligned} N_i &= \begin{pmatrix} p \\ n \end{pmatrix}, & \pi_i^j &= \begin{pmatrix} \pi^0/\sqrt{2} & \pi^+ \\ \pi^- & -\pi^0/\sqrt{2} \end{pmatrix}, \\ \bar{N}^i &= (\bar{p}, \bar{n}), & \bar{\pi}_i^j &= \begin{pmatrix} \bar{\pi}^0/\sqrt{2} & \bar{\pi}^- \\ \bar{\pi}^+ & -\bar{\pi}^0/\sqrt{2} \end{pmatrix}. \end{aligned} \quad (136.20)$$

One can form from them the two independent isoscalars

$$\bar{N}^i N_i \bar{\pi}_j^k \pi_k^j \quad \text{and} \quad \bar{N}^i \bar{\pi}_i^j \pi_j^k N_k$$

so that the amplitude is written in the form of a linear combination

$$f = f_1 (\bar{N}^i N_i \bar{\pi}_j^k \pi_k^j) + f_2 (\bar{N}^i \bar{\pi}_i^j \pi_j^k N_k). \quad (136.21)$$

From (136.20) we have for the isospin part of the wave functions of the particles of the initial and final states

$$\begin{aligned} \text{for the proton} & \quad p \rightarrow N_1 = 1, \quad \bar{p} \rightarrow \bar{N}^1 = 1 \\ \text{for the neutron} & \quad n \rightarrow N_2 = 1, \quad \bar{n} \rightarrow \bar{N}^2 = 1 \\ \text{for the } \pi^+ \text{-meson} & \quad \pi^+ \rightarrow \pi_1^2 = 1, \quad \bar{\pi}^+ \rightarrow \bar{\pi}_2^1 = 1 \\ \text{for the } \pi^- \text{-meson} & \quad \pi^- \rightarrow \pi_2^1 = 1, \quad \bar{\pi}^- \rightarrow \bar{\pi}_1^2 = 1 \\ \text{for the } \pi^0 \text{-meson} & \quad \pi^0 \rightarrow \pi_1^1 = \frac{1}{\sqrt{2}}, \pi_2^2 = -\frac{1}{\sqrt{2}}, \bar{\pi}^0 \rightarrow \bar{\pi}_1^1 = \frac{1}{\sqrt{2}}, \bar{\pi}_2^2 = -\frac{1}{\sqrt{2}} \end{aligned}$$

(all other components are equal to zero).

Substituting these wave functions into the amplitude (136.21), we obtain

$$f^{(+)} = f_1, \quad f^{(-)} = f_1 + f_2, \quad f^{(0)} = -f_2/\sqrt{2}, \quad (136.22)$$

where f_1 and f_2 are functions of spatial-spin variables, unknown but the same for all pion-nucleon processes. From (136.22) there follows, in particular, relation (136.15), and therefore also the triangle inequalities. It is known from experiment that in the region of high energies and small angles the cross section of the charge-exchange process $\pi^- + p \rightarrow \pi^0 + n$ is small compared to

elastic cross sections. Hence assuming $f_2 \approx 0$ we obtain the approximate equality of the differential cross sections in the forward direction for the elastic scattering of π^+ and π^- -mesons by protons:

$$\left. \frac{d\sigma^{(+)}}{d\Omega} \right|_{\theta \rightarrow 0} \approx \left. \frac{d\sigma^{(-)}}{d\Omega} \right|_{\theta \rightarrow 0}. \quad (136.23)$$

The amplitudes f_i can be expressed in terms of the amplitudes $f^{(T)}$ and vice versa. For this it suffices to compare relations (136.22) and (136.14), hence

$$f_1 = f^{(\frac{3}{2})}; \quad f_2 = -\frac{2}{3}f^{(\frac{3}{2})} + \frac{2}{3}f^{(\frac{1}{2})}. \quad (136.24)$$

In concluding this section we indicate a simple method of determining the number of independent invariant amplitudes. According to the rule of addition of angular momenta we find possible values of the isospin T of the initial and final states. By virtue of charge independence only transitions with conservation of T are possible, hence the number of independent amplitudes is defined by the number of values of the isospin which occur in both the initial and final states. In our example the isospin of the initial and final states are equal to $\frac{3}{2}$ and $\frac{1}{2}$, for which there are two amplitudes: $f^{(\frac{3}{2})}$ and $f^{(\frac{1}{2})}$, or f_1 and f_2 .

§137. The unitary group SU(3) and its representations

In the preceding sections we studied some consequences of the isospin invariance of the strong interaction based on the group SU(2) and one most suitable for the description of the symmetry properties of nucleons and π -mesons. However, after the discovery of strange particles the framework of this group turned out to be too narrow, because its rank is equal to 1 and it gives only one conserved additive quantum number (the isospin component T_3) by means of which the terms of a given isomultiplet are classified. By an additive quantum number is meant a quantity whose value for a certain system is equal to the sum of its values for the subsystems. In this sense, T is not an additive characteristic. At the same time there is at least one more characteristic of this type, the hypercharge Y (or strangeness), so that it is natural to try to group several isomultiplets with different hypercharges into one supermultiplet. For this a group of rank 2 is necessary, the mutually commuting generators of which give two simultaneously measurable conserved quantum numbers characterizing the terms of the supermultiplet, so that they can be identified with T_3 and Y . On the other hand, from the requirement that the new theory contain the results of the old it follows that

the isogroup must be a part (subgroup) of a new larger symmetry group. These requirements can most simply and naturally be satisfied if one postulates the approximate invariance of the strong interaction with respect to the group $SU(3)$ which we shall henceforth call unitary (in the narrow sense of the word)*. The corresponding mathematical apparatus is very close to the formalism presented in §132 the results of which we shall frequently refer to.

The group $SU(3)$ is understood to be the set of all unitary and unimodular matrices of third order which correspond to linear transformations in a three-dimensional complex space conserving the quadratic form $x^\dagger x = x^i x_i = x_1^* x_1 + x_2^* x_2 + x_3^* x_3$ (the indices i, j and so on now run over the values 1, 2, 3). The generators of this group are Hermitian square 3×3 matrices λ_α with trace zero:

$$\lambda_\alpha^\dagger = \lambda_\alpha, \quad \text{Tr } \lambda_\alpha = 0. \quad (137.1)$$

Nine conditions of hermiticity and 1 condition of equality to zero of the trace are imposed upon 18 real parameters of the complex 3×3 matrices, so that there are 8 independent matrices with the properties (137.1), which defines the dimension of the group $SU(3)$. Among the matrices λ_α there are two mutually commuting ones (the rank of the group $SU(3)$ is equal to 2) which can simultaneously be diagonalized. We choose the representation in which λ_3 and λ_8 are diagonal:

$$\begin{aligned} \lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \\ \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned} \quad (137.2)$$

Hence it follows that if all parameters ω_α of the group $SU(3)$, except for the first three, are set equal to zero, then we obtain the isogroup $SU(2)$.

The representations of the group $SU(3)$ are constructed in a completely analogous way to §132. However, the mutually conjugated representations

* It should be noted that in 1961–1964 the situation was not completely clear. since it was impossible to make an unambiguous choice between $SU(3)$ and the so-called group G_2 ; some preference even was given to the latter. The problem was finally solved in 1964 when the Ω -hyperon was discovered (see §138).

will now not be equivalent, so that the given irreducible representation (p, q) is defined by two numbers, p and q (rather than one number $p+q$). The definitions and relations (132.13)–(132.16) remain valid as before if τ_α is replaced in them by λ_α and if it is assumed that the Roman indices run over values from 1 to 3, and Greek indices from 1 to 8.

Let us find the number of independent components of the matrix $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ which is symmetric separately in all subscripts and superscripts and which has zero trace over any pair of superscript and subscript, i.e. let us determine the dimension $N(p, q)$ of the irreducible representation (p, q) . By virtue of symmetry only components differing in one of the numbers p_1, p_2, p_3 (the numbers of ones, twos and threes among the subscripts) are different. Since $p_1 + p_2 + p_3 = p$, then for a given p_1 the number p_2 can vary from 0 up to $p - p_1$, which gives $p - p_1 + 1$ different components. Now varying p_1 from 0 to p , we obtain the total number of different components for fixed superscripts:

$$N(p) = \sum_{p_1=0}^p (p - p_1 + 1) = \frac{1}{2}(p+1)(p+2).$$

Analogously, the number of different components for fixed subscripts is equal to $N(q) = \frac{1}{2}(q+1)(q+2)$, there are in all $N(p)N(q)$ components. But they are still not independent because they are related by the conditions that the traces are equal to zero. By virtue of symmetry it is sufficient that the trace with respect to one of the pairs of superscript and subscript reduce to zero; this trace will be a matrix with $p-1$ subscripts and $q-1$ superscripts, having thus $N(p-1)N(q-1)$ components. Thus $N(p, q) = N(p)N(q) - N(p-1)N(q-1)$ and, finally,

$$N(p, q) = \frac{1}{2}(p+1)(q+1)(p+q+2). \quad (137.3)$$

Let us enumerate the most important representations of the group SU(3):

1. φ $(0, 0) \equiv 1$ unitary scalar or singlet ($N=1$),
2. φ_i and φ^i $(1, 0) \equiv 3$ and $(0, 1) \equiv \bar{3}$ unitary spinors or triplets ($N=3$),
3. φ_{ij} and φ^{ij} $(2, 0) \equiv 6$ and $(0, 2) \equiv \bar{6}$ sextets ($N=6$),
4. φ_i^j $(1, 1) \equiv 8$ unitary vector or octet ($N=8$),
5. φ_{ijk} and φ^{ijk} $(3, 0) \equiv 10$ and $(0, 3) \equiv \bar{10}$ decuplets ($N=10$),
6. φ_{ij}^k and φ_i^{jk} $(2, 1) \equiv 15$ and $(1, 2) \equiv \bar{15}$ 15-plets ($N=15$),
7. φ_{ij}^k $(2, 2) \equiv 27$ ($N=27$),

and so on. A very convenient notation is given here, which immediately

indicates the dimension of the irreducible representation. Attention should be drawn, for example, to the representations (2,0) and (1,1); the corresponding wave functions have the same number of indices (namely 2), but the dimensions of the multiplets are different (6 and 8 respectively). Such a situation cannot be encountered in the group SU(2) by virtue of the equivalence of its mutually conjugate representations.

Let us now write down the decompositions of some direct products of representations into irreducible representations:

$$(1,0) \otimes (0,1) = (0,0) \oplus (1,1) \quad (137.4a)$$

$$\text{or} \quad 3 \otimes \bar{3} = 1 \oplus 8,$$

$$(1,0) \otimes (1,0) \otimes (1,0) = (0,0) \oplus (1,1) \oplus (1,1) \oplus (3,0) \quad (137.4b)$$

$$\text{or} \quad 3 \otimes 3 \otimes 3 = 1 \oplus 8 \oplus 8 \oplus 10,$$

$$(1,0) \otimes (1,0) \otimes (0,1) = (1,0) \oplus (1,0) \oplus (0,2) \oplus (2,1) \quad (137.4c)$$

$$\text{or} \quad 3 \otimes 3 \otimes \bar{3} = 3 \oplus 3 \oplus \bar{6} \oplus 15,$$

$$(1,1) \otimes (1,1) = (0,0) \oplus (1,1) \oplus (1,1) \oplus (3,0) \oplus (0,3) \oplus (2,2) \quad (137.4d)$$

$$\text{or} \quad 8 \otimes 8 = 1 \oplus 8 \oplus 8 \oplus 10 \oplus \bar{10} \oplus 27,$$

$$(1,1) \otimes (3,0) = (1,1) \oplus (3,0) \oplus (2,2) \oplus (3,1) \quad (137.4e)$$

$$\text{or} \quad 8 \otimes 10 = 8 \oplus 10 \oplus 27 \oplus 35.$$

Let us prove, for example, decompositions (137.4a) and (137.4d). In the first case the function $\varphi_i \chi^j$ contains the non-zero trace $\varphi_i \chi^i$ which is a scalar (representation (0,0)); the function $\varphi_i \chi^j - \frac{1}{3} \delta_i^j \varphi_k \chi^k$, which remains after the separation of the non-zero trace, transforms according to the irreducible representation (1,1). The proof of formula (137.4d) is somewhat more complex. First of all, if we symmetrize the function $\varphi_i^j \chi_k^l$ with respect to the subscripts and superscripts and separate non-zero traces, then we obtain representation (2,2). Further, two different traces with respect to only one of the pairs of indices (we recall that $\varphi_i^i = \chi_i^i = 0$) after the separation of the non-zero traces with respect to the remaining pair of indices, i.e. the two functions

$$\varphi_i^j \chi_k^i - \frac{1}{3} \delta_k^j \varphi_i^l \chi_l^i \quad \text{and} \quad \varphi_i^j \chi_j^k - \frac{1}{3} \delta_i^k \varphi^l \chi_l^j$$

transform according to the two representations (1,1) and (1,1)'. The total trace $\varphi_i^i \chi_j^j$ is a unitary scalar, so that we have the representation (0,0). Finally, on separating symmetric parts there arise two functions of the type

$$\psi_{[ik]}^{[jl]} \equiv \varphi_i^j \chi_k^l + \varphi_k^j \chi_i^l - \varphi_i^l \chi_k^j - \varphi_k^l \chi_i^j \quad \text{and} \quad \psi_{[ik]}^{[jl]} \equiv \varphi_i^j \chi_k^l + \varphi_i^l \chi_k^j - \varphi_k^j \chi_i^l - \varphi_k^l \chi_i^j$$

containing 10 independent components each. It can be shown that the lower pair $\{ik\}$ is equivalent to one upper index and vice versa, hence we obtain the representations (3,0) and (0,3). We can convince ourselves of the validity of this decomposition by comparing the dimensions on the left-hand and right-hand sides of formula (137.4d): $8 \times 8 = 1+8+8+10+10+27 = 64$. Other formulae (137.4) are proved in an analogous way.

The group SU(3) has two invariant operators whose eigenvalues serve for the classification of irreducible representations. But we have already unambiguously characterized the representations also by two numbers p and q . Hence we shall not write down and analyse the invariant operators; we shall not need them in what follows.

For the classification of the basis elements of a multiplet, which can be chosen to be definite combinations of the matrices $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ with only one non-zero element, we shall make use of the eigenvalues of the diagonal generators Λ_3, Λ_8 . From the explicit expressions of the type (132.15) for the generators Λ_α it follows that they act on the functions $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ with one non-zero element in the following way:

$$\Lambda_\alpha \varphi_{i_1 \dots i_p}^{j_1 \dots j_q} = \sum_{s=1}^p (\lambda_\alpha)_{i_s}^{i'_s} \varphi_{i_1 \dots i_{s-1} i'_s i_{s+1} \dots i_p}^{j_1 \dots j_q} - \sum_{s=1}^q (\lambda_\alpha)_{j'_s}^{j_s} \varphi_{i_1 \dots i_p}^{j_1 \dots j_{s-1} j'_s j_{s+1} \dots j_q}. \quad (137.5)$$

Let there be among p subscripts, p_3 indices 3 and $p-p_3$ indices 1 and 2, and among q superscripts, q_3 indices 3 and $q-q_3$ indices 1 and 2. Since the matrix λ_8 gives $1/\sqrt{3}$ in application to each superscript and subscript 1, 2, and $-2/\sqrt{3}$ in application to any index 3, then according to (137.5) the eigenvalues of the generators Λ_8 are equal to

$$\Lambda_8 = \sqrt{3} \left[\frac{1}{3}(p-q) - p_3 + q_3 \right]. \quad (137.6)$$

Analogously, for the eigenvalues of the generator Λ_3 we have

$$\Lambda_3 = (p_1 - p_2 - q_1 + q_2). \quad (137.7)$$

If no threes were present among the indices, i.e. if the group SU(2) were considered, then the relations $p_1 + p_2 = p$ and $q_1 + q_2 = q$ would be fulfilled. In this case formula (137.7) would go over into (132.25).

Two quantum numbers Λ_3, Λ_8 are not sufficient, however, for the unambiguous classification of the basis elements of a given multiplet. They are complemented, say, by the eigenvalue of the operator $\Lambda^2 = \Lambda_1^2 + \Lambda_2^2 + \Lambda_3^2$. From (137.2) it follows that

$$\lambda^2 \equiv \lambda_1^2 + \lambda_2^2 + \lambda_3^2 = 3 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (137.8)$$

and hence the matrix λ^2 does not commute with all the matrices λ_α . By virtue of the fact that the expressions for the generators are completely analogous to those given in §132 (see (132.15)) this statement is valid also for Λ^2 . For this reason the situation with the eigenvalues of the operator Λ^2 is somewhat more complex, since there may correspond several eigenvalues of the operator Λ^2 to the component $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ referring to the eigenvalues $\Lambda_3 = \Lambda_8 = 0$, i.e. a peculiar degeneracy occurs. It is easily shown that

$$\Lambda^2 \varphi_{i_1 \dots i_p}^{j_1 \dots j_q} = (p_1 + p_2 + q_1 + q_2)(p_1 + p_2 + q_1 + q_2 + 2) \varphi_{i_1 \dots i_p}^{j_1 \dots j_q}, \quad (137.9)$$

but this statement will be valid generally only if $\Lambda_3 \neq 0$ or $\Lambda_8 \neq 0$. In the case $\Lambda_3 = \Lambda_8 = 0$ the factor in the right-hand side of (137.9) gives, however, the maximum eigenvalue of the operator Λ^2 .

§138. The eightfold way formalism and unitary multiplets

Suppose that in nature there were a superstrong interaction of elementary particles invariant with respect to the group $SU(3)$. Then the wave functions of hadrons would transform according to some of its irreducible representations, i.e. they are the products of ordinary spatial-spin wave functions with the unitary matrices $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$. As a result all hadrons would be distributed in unitary multiplets, characterized by a pair of numbers p and q , spin, parity, baryonic number and other quantities not associated with the group $SU(3)$ (see §133). Individual hadrons within a multiplet are classified by the eigenvalues of the generators Λ_3 and Λ_8 and of the operator Λ^2 , which will be identified below with T_3 , Y and T^2 .

If it is assumed that there is only the superstrong interaction, then these quantum numbers must be conserved, the transitions occurring as a result of the reactions being possible only within one unitary multiplet, as in the case of isospin invariance. Furthermore, quantum-mechanical degeneracy must result, and the masses of the particles contained in one unitary multiplet must

all be strictly the same. On the other hand, experiment shows that the masses of the particles having different hypercharge (for the same spin, parity and baryonic number) differ sharply: for example, the mass difference between the nucleon and Ξ -hyperon amounts to about 30% of the mass of the latter. This means that the ordinary strong interaction must already essentially violate the SU(3) invariance, and to a much larger degree than the electromagnetic interaction violates the isospin symmetry. But in the strong interaction Y , T and T_3 are still conserved, and this allows one to obtain the transformation properties of the Hamiltonian of the interaction violating the symmetry with respect to unitary transformations from the group SU(3). This in its turn makes it possible to obtain definite relations between the masses of the particles contained in a given unitary multiplet.

Instead of the generators Λ_1 , Λ_2 , Λ_3 and Λ_8 it is convenient to introduce the operators

$$\hat{T}_1 = \frac{1}{2}\Lambda_1, \quad \hat{T}_2 = \frac{1}{2}\Lambda_2, \quad \hat{T}_3 = \frac{1}{2}\Lambda_3 \quad (138.1)$$

and

$$\hat{Y} = \frac{1}{\sqrt{3}}\Lambda_8. \quad (138.2)$$

It is natural to identify the operators \hat{T}_1 , \hat{T}_2 , \hat{T}_3 with the operators of the isospin components, and the operator

$$\hat{T}^2 \equiv \hat{T}_1^2 + \hat{T}_2^2 + \hat{T}_3^2 = \frac{1}{4}\Lambda^2 \quad (138.3)$$

with the operator of the isospin squared. From (137.7) and (137.9) it follows that the eigenvalues of \hat{T}_3 and \hat{T}^2 are equal to (see however the remark at the end of § 137)

$$T_3 = \frac{1}{2}(p_1 - p_2 - q_1 + q_2) \quad (138.4)$$

and

$$T(T+1) = \frac{1}{4}(p_1 + p_2 + q_1 + q_2)(p_1 + p_2 + q_1 + q_2 + 2). \quad (138.5)$$

The operator $\hat{Y} = \Lambda_8/\sqrt{3}$ can be identified with the hypercharge operator, so that, according to (137.6),

$$Y = \frac{1}{3}(p - q) - p_3 + q_3. \quad (138.6)$$

Such an identification is not unambiguous and corresponds to the so-called eightfold formalism or eightfold way proposed in 1961 by Gell-Mann and independently by Ne'eman. Its suitability is justified a posteriori, since within the framework of the eightfold way the real elementary particles are described in the best of all possible ways. Another choice of the operator \hat{Y} is considered in § 141.

From (138.6) and the requirement that the hypercharge has an integer value it follows that it is necessary to consider only those representations of the group SU(3) for which the difference $p-q$ is a multiple of 3

$$p - q = 3n \quad (138.7)$$

i.e. the representations

$$(0,0) = 1, \quad (1,1) = 8, \quad (3,0) = 10, \quad (0,3) = \bar{10}, \quad (2,2) = 27, \quad (138.8)$$

and so on. In view of the importance of the representation $(1,1) = 8$ in the approach considered, it is called the eightfold formalism. Finally, we note that in correspondence with the Gell-Mann–Nishijima relation (§ 129, 133) we assume by definition that the charge operator is

$$\hat{Q} = \hat{T}_3 + \frac{1}{2} \hat{Y} = \frac{1}{2} \Lambda_3 + \frac{1}{2\sqrt{3}} \Lambda_8. \quad (138.9)$$

Let us now consider concrete unitary multiplets of hadrons. We turn first of all to the octet with the wave function φ_i^j , and investigate its content in isospin T , its component T_3 , hypercharge Y and electric charge Q . Making use of relations (138.3)–(138.6) and (138.9), one can immediately draw up table 5. In the last three columns are shown the stable particles, pseudoscalar and vector mesons and baryons with spin $\frac{1}{2}$, which have the corresponding quan-

Table 5

Component	T	T_3	Y	Q	Meson 0^-	Meson 1^-	Baryon
φ_1^1	1,0*	0	0	0	π^0, η	$\rho^0; \omega, \varphi$	Σ^0, Λ
φ_2^2	1,0*	0	0	0	π^0, η	$\rho^0; \omega, \varphi$	Σ^0, Λ
φ_1^2	1	+1	0	+1	π^+	ρ^+	Σ^+
φ_2^1	1	-1	0	-1	π^-	ρ^-	Σ^-
φ_1^3	$\frac{1}{2}$	$+\frac{1}{2}$	+1	+1	K^+	K^{*+}	p
φ_2^3	$\frac{1}{2}$	$-\frac{1}{2}$	+1	0	K^0	K^{*0}	n
φ_3^1	$\frac{1}{2}$	$-\frac{1}{2}$	-1	-1	\bar{K}^+	\bar{K}^{*+}	Ξ^-
φ_3^2	$\frac{1}{2}$	$+\frac{1}{2}$	-1	0	\bar{K}^0	\bar{K}^{*0}	Ξ^0
φ_3^3	0	0	0	0	η	ω, φ	Λ

tum numbers. It is remarkable that all stable pseudoscalar mesons, whose number is just 8, and all 8 stable baryons (except for Ω^- whose spin is equal to $\frac{3}{2}$) are involved here. Hence it is natural to assume that these two groups of particles just form unitary octets. The matrices of their wave functions have the following form:

$$B_i^j = \begin{pmatrix} \Sigma^0/\sqrt{2} + \Lambda/\sqrt{6} & \Sigma^+ & p \\ \Sigma^- & -\Sigma^0/\sqrt{2} + \Lambda/\sqrt{6} & n \\ \Xi^- & \Xi^0 & -2\Lambda/\sqrt{6} \end{pmatrix}; \quad (138.10)$$

$$M_i^j = \begin{pmatrix} \pi^0/\sqrt{2} + \eta/\sqrt{6} & \pi^+ & K^+ \\ \pi^- & -\pi^0/\sqrt{2} + \eta/\sqrt{6} & K^0 \\ \bar{K}^+ & \bar{K}^0 & -2\eta/\sqrt{6} \end{pmatrix}.$$

The choice of just such coefficients is dictated by the requirement that the trace of the matrix be equal to zero and by the normalization considerations (see §133). The meson octet contains particles together with antiparticles. This is accounted for by the fact that there now remain no quantum numbers which are not involved in the group SU(3) by means of which one could distinguish, say, between K^+ and \bar{K}^+ (see §133). For baryons, however, such a number exists (the baryonic number B), so that the corresponding antiparticles form an independent octet:

$$\bar{B}_i^j = \begin{pmatrix} \bar{\Sigma}^0/\sqrt{2} + \bar{\Lambda}/\sqrt{6} & \bar{\Sigma}^- & \bar{\Xi}^- \\ \bar{\Sigma}^+ & -\bar{\Sigma}^0/\sqrt{2} + \bar{\Lambda}/\sqrt{6} & \bar{\Xi}^0 \\ \bar{p} & \bar{n} & -2\bar{\Lambda}/\sqrt{6} \end{pmatrix}. \quad (138.11)$$

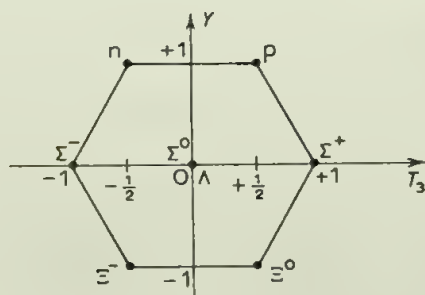


Fig. V.39

Unitary multiplets are conveniently presented in diagrams called weight diagrams. In this case an orthogonal reference frame is chosen, the hypercharge being plotted on an axis and the isospin component on the other axis. Thus, for example, for the baryon octet we have fig. V.39. The situation with vector mesons is somewhat more complex, because at a place analogous to that occupied by the η -meson or Λ -hyperon there are two pretenders, ω and φ . We shall discuss it at the end of this section.

Let us now draw up the table of quantum numbers of the particles which can be contained in the decuplet whose wave function is $\varphi_{[ijk]}$, table 6.

Table 6

Component	T	T_3	Y	Q	Baryon $\frac{3}{2}^+$
φ_{111}	$\frac{3}{2}$	$+\frac{3}{2}$	+1	+2	Δ_{1236}^{++}
φ_{112}	$\frac{3}{2}$	$+\frac{1}{2}$	+1	+1	Δ_{1236}^+
φ_{122}	$\frac{3}{2}$	$-\frac{1}{2}$	+1	0	Δ_{1236}^0
φ_{222}	$\frac{3}{2}$	$-\frac{3}{2}$	+1	-1	Δ_{1236}^-
φ_{113}	1	+1	0	+1	Σ_{1385}^+
φ_{123}	1	0	0	0	Σ_{1385}^0
φ_{223}	1	-1	0	-1	Σ_{1385}^-
φ_{133}	$\frac{1}{2}$	$+\frac{1}{2}$	-1	0	Ξ_{1530}^0
φ_{233}	$\frac{1}{2}$	$-\frac{1}{2}$	-1	-1	Ξ_{1530}^-
φ_{333}	0	0	-2	-1	?

At the time when an analogous table was drawn up (1961–1962) the place at which the question mark is standing was unoccupied, because no particle with hypercharge -2 was known. Thus the eightfold way predicted the existence of a new hyperon with spin $\frac{3}{2}^+$ and with the quantum numbers indicated in the table. Moreover, the mass of this particle (about 1680 MeV) was also known approximately (see §140). At the beginning of 1964 such a particle was indeed discovered: this is the rather stable Ω^- -hyperon (see the table of elementary particles given in §129). This fact eliminated any doubt as to the validity of the eightfold way and, in general, of unitary symmetry, which is now as classical as isospin symmetry.

The weight diagram for the decuplet is given in fig. V.40.

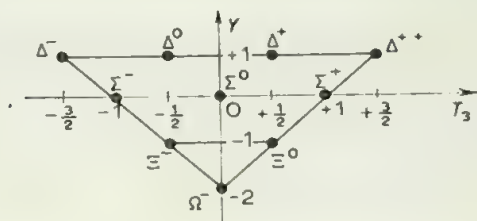


Fig. V.40.

Formula (138.7) also allows for the existence of unitary singlets. Since in this case all generators reduce to zero, then for a unitary scalar particle $T = T_3 = Y = Q = 0$. However, among the multitude of resonance states there is up to now none whose wave function could be quite trustworthily considered as a unitary scalar.

That is why unitary singlets play a decisive role in resolving the difficulty mentioned above with vector mesons. It can be assumed that real ω - and ϕ -mesons represent different superpositions of the unitary singlet state ϕ' and the octet state ω' analogous to the η -meson. In the case of strict unitary symmetry such a mixing of components of multiplets of different nature is forbidden, but as a result of the violation of the symmetry by the real strong interaction it is no longer impossible. Thus vector mesons form a nonet and not an octet. Apropos of this we shall confine ourselves only to the remarks of general character made above, referring the reader to the literature.

Let us now enumerate, without comment, some unitary multiplets in which resonances are distributed. It turns out that there exists a family of 9 mesons 2^+ (it contains, in particular, the f^0 -meson mentioned in §129), which also form a nonet. The nonet of mesons 1^+ , octet of baryon resonances $\frac{3}{2}^-$ and octets of baryon resonances $\frac{5}{2}^+$ and $\frac{7}{2}^+$ are somewhat more doubtful. At present insufficient data are available for a final solution of the problem.

§139. Some consequences of strict unitary symmetry

In this section some physical consequences of the hypothesis of strict unitary symmetry for hadrons are described briefly. It should be stressed immediately that they cannot pretend to be in good agreement with experimental data, since the real strong interaction already violates to a considerable degree the $SU(3)$ invariance of the theory. A more realistic scheme is outlined in the next section.

Let us consider first of all the interaction of stable baryons of spin $\frac{1}{2}^+$ (octet B_i^j) with pseudoscalar mesons (octet M_i^j). In quantum field theory the baryon and meson functions are considered to be operators in the space of occupation numbers. Analogously to what was done in §135, the interaction Hamiltonian density is to be constructed in the form of an invariant combination of the functions \bar{B}_i^j, B_k^l and M_m^n . We first of all form unitary scalars from these functions. Noting that the two octets

$$\bar{B}_i^j B_j^k - \frac{1}{3} \delta_i^k \bar{B}_i^j B_j^m \quad \text{and} \quad \bar{B}_j^k B_i^l - \frac{1}{3} \delta_i^k \bar{B}_j^l B_j^m$$

can be constructed from \bar{B}_i^j and B_k^l , we contract these matrices with the meson function M_k^l . Taking into account that $\delta_i^k M_k^l = M_i^l = 0$ we arrive at the two unitary scalars $\bar{B}_i^j B_j^k M_k^l$ and $\bar{B}_j^k B_i^l M_k^l$ which exhaust the possible invariants. Usually their sum and difference are chosen, so that the baryon-meson interaction is described by the following Hamiltonian density:

$$\begin{aligned} \hat{H}_{\text{int}} = & \frac{1}{\sqrt{2}} g^{(F)} [\bar{B}_i^j \gamma_5 B_j^k - \bar{B}_j^k \gamma_5 B_i^l] M_k^l + \\ & + \frac{1}{\sqrt{2}} g^{(D)} [\bar{B}_i^j \gamma_5 B_j^k + \bar{B}_j^k \gamma_5 B_i^l] M_k^l \end{aligned} \quad (139.1)$$

(the matrix γ_5 is introduced on the basis of the same considerations as in §135) containing two independent coupling constants $g^{(F)}$ and $g^{(D)}$. For definite reasons the first term of the Hamiltonian (139.1) is called F-coupling, and the second D-coupling.

Making use of (138.10) and (138.11) we write the components of the baryon and meson matrices in terms of the wave functions of isomultiplets:

$$\begin{aligned} B_{33} &= -\frac{2}{\sqrt{6}} \Lambda, & B_b^a &= \Sigma_b^a + \frac{1}{\sqrt{6}} \Lambda \delta_b^a, & B_a^3 &= N_a, & B_3^a &= \Xi^a, \\ \bar{B}_{33} &= -\frac{2}{\sqrt{6}} \bar{\Lambda}, & \bar{B}_b^a &= \bar{\Sigma}_b^a + \frac{1}{\sqrt{6}} \bar{\Lambda} \delta_b^a, & \bar{B}_a^3 &= \bar{\Xi}_a, & \bar{B}_3^a &= \bar{N}^a, \\ M_3^3 &= -\frac{2}{\sqrt{6}} \eta, & M_b^a &= \pi_b^a + \frac{1}{\sqrt{6}} \eta \delta_b^a, & M_a^3 &= K_a, & M_3^a &= \bar{K}^a, \end{aligned}$$

($a, b = 1, 2$ are isospin indices). Substituting them into (139.1), we obtain

$$\begin{aligned}
\hat{H}_{\text{int}} = & -\sqrt{2}g^{(F)}\pi_a^c\bar{\Sigma}_c^b\gamma_5\Sigma_b^a + \frac{1}{\sqrt{2}}(g^{(D)}+g^{(F)})\pi_a^c\bar{N}^a\gamma_5N_c + \\
& + \frac{1}{\sqrt{2}}(g^{(D)}-g^{(F)})\pi_a^c\bar{\Xi}_c\gamma_5\Xi^a + \frac{1}{\sqrt{3}}g^{(D)}\pi_a^c[\bar{\Lambda}\gamma_5\Sigma_c^a + \bar{\Sigma}_c^a\gamma_5\Lambda] + \\
& + \sqrt{\frac{2}{3}}g^{(D)}\eta\bar{\Sigma}_b^a\gamma_5\Sigma_a^b + \frac{1}{2\sqrt{3}}(3g^{(F)}-g^{(D)})\eta\bar{N}^a\gamma_5N_a - \\
& - \frac{1}{2\sqrt{3}}(3g^{(F)}+g^{(D)})\eta\bar{\Xi}_a\gamma_5\Xi^a - \sqrt{\frac{2}{3}}g^{(D)}\eta\bar{\Lambda}\gamma_5\Lambda + \\
& + \frac{1}{\sqrt{2}}(g^{(D)}+g^{(F)})K_a\bar{\Sigma}_b^a\gamma_5\Xi^b + \frac{1}{\sqrt{2}}(g^{(D)}+g^{(F)})\bar{K}^a\bar{\Xi}_b\gamma_5\Sigma_a^b + \\
& + \frac{1}{\sqrt{2}}(g^{(D)}-g^{(F)})K_a\bar{N}^b\gamma_5\Sigma_b^a + \frac{1}{\sqrt{2}}(g^{(D)}-g^{(F)})\bar{K}^a\bar{\Sigma}_a^b\gamma_5N_b + \\
& + \frac{1}{2\sqrt{3}}(3g^{(F)}-g^{(D)})K_a\bar{\Lambda}\gamma_5\Xi^a + \frac{1}{2\sqrt{3}}(3g^{(F)}-g^{(D)})\bar{K}^a\bar{\Xi}_a\gamma_5\Lambda - \\
& - \frac{1}{2\sqrt{3}}(3g^{(F)}+g^{(D)})K_a\bar{N}^a\gamma_5\Lambda - \frac{1}{2\sqrt{3}}(3g^{(F)}+g^{(D)})\bar{K}^a\bar{\Lambda}\gamma_5N_a. \quad (139.2)
\end{aligned}$$

Thus 12 constants of the couplings $\Sigma\Sigma\pi$, $NN\pi$, $\Sigma\Lambda\pi$, $\Xi\Xi\pi$, $\Sigma\Sigma\eta$, $NN\eta$, $\Xi\Xi\eta$, $\Lambda\Lambda\eta$, $\Sigma\Xi K$, $N\Sigma K$, $\Lambda\Xi K$ and $N\Lambda K$ are expressed in terms of only two parameters $g^{(F)}$ and $g^{(D)}$. Each term contained in (139.2) is invariant with respect to isospin transformations, and can be written in explicit form (see §135), as a result of which 64 terms arise, each corresponding to the interaction of actual baryons with a meson.

Let us now consider how the relations between the cross sections for different processes are obtained in unitary invariant theory. Let there be several reactions

$$a_i + b_i \rightarrow c_i + d_i,$$

the particles a_i , b_i , c_i and d_i belonging respectively to the unitary multiplets $(p^{(a)}, q^{(a)})$, $(p^{(b)}, q^{(b)})$, $(p^{(c)}, q^{(c)})$ and $(p^{(d)}, q^{(d)})$. In accordance with the general scheme described in §136, it is necessary in order to obtain the relations between the amplitudes of the reactions mentioned to proceed in the following way.

1. We decompose the direct products of the representations

$$(p^{(a)}, q^{(a)}) \otimes (p^{(b)}, q^{(b)}) \quad \text{and} \quad (p^{(c)}, q^{(c)}) \otimes (p^{(d)}, q^{(d)})$$

into irreducible representations.

2. We write down the independent amplitudes corresponding to the transition from a certain multiplet of the first decomposition to the same multiplet of the second decomposition.

3. Making use of the table of Clebsch–Gordan coefficients of the group $SU(3)$ we expand the wave functions $|a_i b_i\rangle$ and $|c_i d_i\rangle$ of the initial and final states in terms of the basis functions of the multiplets involved in the corresponding expansions. These functions are defined by the type of representation and by the eigenvalues of the isospin operator, isospin component operator and hypercharge operator, so that they should be written in the form $|p, q; T, T_3, Y\rangle$.

4. We substitute the expansions of wave functions into the matrix element of the transition $\langle c_i d_i | a_i b_i \rangle$, as a result of which it turns out to be expressed in terms of a relatively small number of matrix elements of transitions between multiplets of the same type as in item 2.

As an example let us consider the scattering of pseudoscalar mesons by stable baryons of spin $\frac{1}{2}$ (it can easily be calculated that there are in all 27 such processes). In this case all the particles of both initial and final states belong to octets. Making use of the decomposition (137.4d) we conclude that there are 8 independent amplitudes corresponding to the transitions

$$1 \rightarrow 1, \quad 10 \rightarrow 10, \quad \overline{10} \rightarrow \overline{10}, \quad 27 \rightarrow 27, \\ 8' \rightarrow 8', \quad 8 \rightarrow 8, \quad 8 \rightarrow 8', \quad 8' \rightarrow 8.$$

However, making use of the invariance of theory with respect to time reversal, it can be shown that the last two amplitudes are expressed in terms of each other, so that there are 7 independent amplitudes in terms of which the 27 amplitudes of real processes are expressed.

In the case of the scattering of pseudoscalar mesons by $\frac{1}{2}^+$ baryons with the production of a pseudoscalar meson and a baryon resonance $\frac{3}{2}^+$, for example

$$\pi^+ + p \rightarrow \eta + \Delta^{++}$$

we have respectively for the initial and final states the decomposition (137.4d) and (137.4e), so that the amplitudes of the real processes are expressed in terms of only 4 independent transition amplitudes

$$8 \rightarrow 8, \quad 8' \rightarrow 8, \quad 10 \rightarrow 10, \quad 27 \rightarrow 27.$$

However, the tables of Clebsch–Gordan coefficients of the group $SU(3)$ are very cumbersome, and in each actual case it is much more convenient to make use of the method of invariant amplitudes described at the end of

§136. Let us again consider the scattering of pseudoscalar mesons by $\frac{1}{2}^+$ baryons. The total scattering amplitude, which in our case must be constructed from the wave functions B_i^j and M_k^l of the initial state and the wave functions \bar{B}_i^j and \bar{M}_k^l of the final state, is a unitary invariant of the most general form. It can be written in the form of the following combination of nine scalar terms:

$$f = f_1(\bar{B}B)(\bar{M}M) + f_2(\bar{B}M)(\bar{M}B) + f_3(\bar{B}\bar{M})(BM) + f_4(\bar{B}B\bar{M}\bar{M}) + f_5(\bar{B}B\bar{M}M) + \\ + f_6(\bar{B}M\bar{M}B) + f_7(\bar{B}\bar{M}MB) + f_8(\bar{B}M\bar{B}\bar{M}) + f_9(\bar{B}M\bar{B}M). \quad (139.3)$$

For brevity, the traces of the products of matrices of the type B and M are here denoted by parentheses, so that, for example,

$$(\bar{B}B)(\bar{M}M) \equiv \bar{B}_i^j B_j^i \bar{M}_k^l M_l^k,$$

$$(\bar{B}M\bar{B}M) \equiv \bar{B}_i^j \bar{M}_k^l B_l^i M_j^k$$

and so on. The quantities f_α involved in the amplitude (139.3) are functions of spatial-spin variables, which are unknown but the same for all meson-baryon processes.

We already know that the processes considered are described by eight independent amplitudes (without taking into account the invariance with respect to time reversal). Hence a relation must exist between the nine invariants involved in (139.3). Indeed, it does:

$$(\bar{B}B)(\bar{M}M) + (\bar{B}M)(\bar{M}B) + (\bar{B}\bar{M})(BM) = \\ = (\bar{B}B\bar{M}\bar{M}) + (\bar{B}B\bar{M}M) + (\bar{B}M\bar{M}B) + (\bar{B}M\bar{M}B) + (\bar{B}M\bar{B}\bar{M}) + (\bar{B}M\bar{B}M) \quad (139.4)$$

as can easily be seen by direct calculation. In general, the establishment of relations of the type (139.4) in practice turns out to be difficult, but this is not important, because if all the 9 terms are considered to be formally independent, the final results will automatically involve the functions f_α in the form of just eight independent combinations.

As is known, under the operation of time reversal every wave function goes over into its complex conjugate, the initial and final states being exchanged*; in our case

$$B_i^j, \bar{B}_i^j, M_k^l, \bar{M}_k^l \rightarrow B_i^j, \bar{B}_i^j, M_k^l, \bar{M}_k^l. \quad (139.5)$$

When such an exchange is made the first seven terms of (139.3) do not change, while the last two go over into each other, so that from the invariance

* See, for example, L.D.Landau and E.M.Lifshitz, *Quantum mechanics* (Pergamon Press, Oxford, 1965).

with respect to time reversal it follows that $f_8 = f_9$; taking into account relation (139.4) they can be set equal to zero and one can operate with only the first seven invariant amplitudes.

Now, by means of a procedure quite analogous to that described at the end of §136 the amplitudes of all the 27 real processes can be expressed in terms of 7 independent functions f_α , for example

$$\begin{aligned} f(\pi^- p \rightarrow K^+ \Sigma^-) &= f_3 \\ f(K^0 p \rightarrow K^+ \Xi^0) &= f_3 \\ f(K^- p \rightarrow K^- p) &= f_1 + f_2 + f_4 + f_6 \\ f(\pi^- p \rightarrow \pi^- p) &= f_1 + f_6 \\ f(K^- p \rightarrow \pi^- \Sigma^+) &= f_2 + f_4 \end{aligned} \quad (139.6)$$

and so on. On eliminating the functions f_α we shall have the relations between the amplitudes of different processes. Thus from (139.6) it follows that

$$f(\pi^- p \rightarrow K^+ \Sigma^-) = f(\bar{K}^0 p \rightarrow K^+ \Xi^0) \quad (139.7)$$

and

$$f(K^- p \rightarrow K^- p) - f(\pi^- p \rightarrow \pi^- p) = f(K^- p \rightarrow \pi^- \Sigma^+) \quad (139.8)$$

and hence for the cross sections we obtain

$$\sigma(\pi^- p \rightarrow K^+ \Sigma^-) = \sigma(\bar{K}^0 p \rightarrow K^+ \Xi^0) \quad (139.9)$$

and

$$\sqrt{\sigma(\pi^- p \rightarrow \pi^- p)} - \sqrt{\sigma(K^- p \rightarrow K^- p)} \leq \sqrt{\sigma(K^- p \rightarrow \pi^- \Sigma^+)}. \quad (139.10)$$

The cross section on the left-hand side of equality (139.9) has a large value, while the cross section on the right-hand side is small, so that in analysing these processes it is necessary to take into account the violation of unitary symmetry by the strong interactions. On the other hand, inequality (139.10) is fulfilled in the entire energy range. Moreover, for large energies, when the cross section $\sigma(K^- p \rightarrow \pi^- \Sigma^+)$ is very small, (139.10) goes over into the equality

$$\sigma(\pi^- p \rightarrow \pi^- p) \approx \sigma(K^- p \rightarrow K^- p) \quad (139.10a)$$

which is in good agreement with experimental data.

It should be noted that the interpretation of the theoretical predictions of strict unitary symmetry and their comparison with experiment is a rather complex problem, since it is necessary to state beforehand the energy range

in which the violation of $SU(3)$ invariance can be disregarded. Furthermore, the cross sections for different processes involve kinematical factors which are different, because they contain the masses of the particles taking part in the reactions. On the other hand, the relations written above assume equality of the masses of the particles contained in the same multiplet. Hence from the relations between amplitudes one actually obtains relations not between cross sections but between their ratios to kinematical factors, so that one has to take from experiment certain 'corrected' values of the cross sections.

§ 140. Some aspects of violated unitary symmetry

We have above frequently stressed that the real strong interaction violates strict unitary symmetry but, if electromagnetism is disregarded, the isospin T and hypercharge Y are still conserved. Therefore the interaction Hamiltonian density cannot be a unitary scalar, but must be an isoscalar ($T=0$) and must correspond to zero hypercharge ($Y=0$). Thus among the components of unitary multiplets one has to find those for which $T=Y=0$. From formula (138.5) for the maximum value of the isospin, it is seen that all the indices of these components must be threes, so that the equalities $p_3 = p$ and $q_3 = q$ are fulfilled. Then on the basis of expression (138.6) for hypercharge we arrive immediately at the condition $p - q = 0$. Thus zero isospin and zero hypercharge are possessed only by the components of symmetric multiplets of the type (p, p) , i.e. of the multiplets $(1, 1) = 8$, $(2, 2) = 27$ and so on, each index of which is equal to 3. Hence the strong interaction Hamiltonian density must have the following general structure

$$\hat{H}_{\text{int}} = g_0 \hat{H}_0 + g_1 \hat{H}_3^3 + g_2 \hat{H}_{33}^{33} + \dots \quad (140.1)$$

As soon as the terms violating unitary symmetry are introduced the earlier quantum-mechanical degeneracy will be removed, i.e. the masses of the particles belonging to the same multiplet must split. For the derivation of mass formulae we introduce the operator \hat{M} whose eigenvalues are equal to the masses of isomultiplets with definite hypercharge (in consequence of isospin invariance they do not depend on the component T_3):

$$\hat{M}|p, q; T, Y\rangle = m|p, q; T, Y\rangle. \quad (140.2)$$

Under the assumption of strict unitary symmetry the mass operator will be an invariant of the group $SU(3)$, and its eigenvalues in a given multiplet will be the same. It is natural to assume that, when unitary symmetry is violated,

\hat{M} acquires a structure analogous to (140.1), it being assumed that $g_2 \ll g_1$; hence

$$\hat{M} = \hat{M}_0 + \hat{M}_3^2. \quad (140.3)$$

It is convenient to introduce the matrix M of the mass operator

$$M = \langle p, q; T', Y' | \hat{M} | p, q; T, Y \rangle \quad (140.4)$$

which from (140.2) is diagonal; its elements define the masses of the terms of the multiplet. This matrix represents a certain bilinear combination of the wave functions $\varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$ of a given multiplet (p, q) and the wave function $\bar{\varphi}_{j'_1 \dots j'_q}^{i'_1 \dots i'_p}$ of the conjugate multiplet (q, p) . From (140.3) it follows that it must contain an invariant part and a term corresponding to the 3-3 component of the octet. From the wave functions mentioned a scalar can be constructed:

$$m_0 \bar{\varphi}_{j_1 \dots j_q}^{i_1 \dots i_p} \varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$$

and two octets

$$a_1 \bar{\varphi}_{j_1 \dots j_q}^{i_1 \dots i_{p-1} j} \varphi_{i_1 \dots i_{p-1} i}^{j_1 \dots j_q} - \frac{1}{3} a_1 \delta_i^j \bar{\varphi}_{j_1 \dots j_q}^{i_1 \dots i_p} \varphi_{i_1 \dots i_p}^{j_1 \dots j_q}$$

and

$$a_2 \bar{\varphi}_{j_1 \dots j_{q-1} i}^{i_1 \dots i_p} \varphi_{i_1 \dots i_p}^{j_1 \dots j_{q-1} j} - \frac{1}{3} a_2 \delta_i^j \bar{\varphi}_{j_1 \dots j_q}^{i_1 \dots i_p} \varphi_{i_1 \dots i_p}^{j_1 \dots j_q},$$

hence

$$M = a_0 \bar{\varphi}_{j_1 \dots j_q}^{i_1 \dots i_p} \varphi_{i_1 \dots i_p}^{j_1 \dots j_q} + a_1 \bar{\varphi}_{j_1 \dots j_q}^{i_1 \dots i_{p-1} j} \varphi_{i_1 \dots i_{p-1} i}^{j_1 \dots j_q} + a_2 \bar{\varphi}_{j_1 \dots j_{q-1} i}^{i_1 \dots i_p} \varphi_{i_1 \dots i_p}^{j_1 \dots j_{q-1} j} \quad (140.5)$$

where $a_0 \equiv m_0 - \frac{1}{3}(a_1 + a_2)$, and m_0 , a_1 and a_2 are certain parameters. Thus in the general case the mass formula contains at the most 3 parameters (for multiplets of the type $(3n, 0)$ and $(0, 3n)$ only two, since in these cases there are only superscripts or subscripts), of which m_0 corresponds to the mass of the members of the multiplet under the assumption of strict unitary symmetry. We note that, according to Feynman, for boson multiplets it is necessary to consider the matrix M^2 instead of the matrix M , since bosons, in contrast to fermions, obey an equation of second order.

For the octet of baryons of spin $\frac{1}{2}^+$ formula (140.5) goes over into

$$M = a_0 \bar{B}_i^j B_i^j + a_1 \bar{B}_3^j B_3^j + a_2 \bar{B}_3^i B_i^3. \quad (140.6)$$

Making use of (138.10) and calculating the matrix elements corresponding to each baryon, we obtain

$$\begin{aligned} m_N &= a_0 + a_2, & m_{\Xi} &= a_0 + a_1, \\ m_{\Sigma} &= a_0, & m_{\Lambda} &= a_0 + \frac{2}{3}(a_1 + a_2). \end{aligned}$$

From this follows the Gell-Mann mass formula

$$3m_{\Lambda} + m_{\Sigma} = 2(m_N + m_{\Xi}) \quad (140.7)$$

which is in good agreement with experiment: the right-hand side adds up to 4518 MeV, while the left-hand side adds up to 4535 MeV, i.e. the accuracy is to within about 0.4%. Taking into account that in the pseudoscalar meson octet there is a K-meson in place of the nucleon, and in place of the Ξ -hyperon there is a \bar{K} -meson, the masses of the K and \bar{K} being equal, we have from an analogue of (140.6)

$$3m_{\eta}^2 + m_{\pi}^2 = 4m_K^2. \quad (140.7')$$

This relation agrees with experiment to within 5%. In the case of vector mesons the situation is made more complex by the presence of ω - φ mixing (see § 138), and we shall not discuss it.

For the decuplet of particles of spin $\frac{3}{2}^+$ formula (140.5) goes over into

$$M = b_0 \bar{B}^{ijk} B_{ijk} + b_1 \bar{B}^{3ij} B_{3ij}. \quad (140.8)$$

Hence, making use of the results of § 138, we obtain

$$\begin{aligned} m_{\Delta} &= b_0, & m_{\Sigma^*} &= b_0 + \frac{1}{3}b_1, \\ m_{\Xi^*} &= b_0 + \frac{2}{3}b_1, & m_{\Omega} &= b_0 + b_1. \end{aligned}$$

From these formulae the interval rule follows:

$$m_{\Sigma^*} - m_{\Delta} = m_{\Xi^*} - m_{\Sigma^*} = m_{\Omega} - m_{\Xi^*}. \quad (140.9)$$

For the first equality we have 147 and 145 MeV (accuracy of the order of 1%), and from the second equality the mass of the Ω -hyperon can be predicted to be 1676 MeV, which was brilliantly confirmed by experiment: $m_{\Omega}^{\text{exp}} = 1675$ MeV.

Okubo has derived a general mass formula valid for all unitary multiplets:

$$m^{2-|B|} = a(p, q) + b(p, q)BY + c(p, q)[T(T+1) - \frac{1}{4}Y^2]. \quad (140.10)$$

In the case of violated SU(3) symmetry the amplitudes of the scattering of one unitary multiplet by another will no longer be invariant; in addition to the scalar part they will contain the 3-3 components of the octet. As an

example we write down the first term of formula (139.3), which now assumes the form

$$f_1 = f_{1,0}(\bar{B}B)(\bar{M}M) + f_{1,1}\bar{B}_i^3 B_3^i(\bar{M}M) + f_{1,2}\bar{B}_3^i B_i^3(\bar{M}M) + \\ + f_{1,3}(\bar{B}B)\bar{M}_i^3 M_3^i + f_{1,4}(\bar{B}B)M_3^i \bar{M}_i^3. \quad (140.11)$$

In view of the enormous number of independent arbitrary parameters arising in such a scheme, its heuristic value decreases sharply, since the physical information that it provides becomes very small.

However, if one adopts the reasonable and experimentally confirmed hypothesis of Okun and Pomeranchuk, that asymptotically at very high energies the cross sections for charge exchange inelastic processes are negligibly small in comparison with the cross sections for ordinary elastic scattering, then in this energy range only the first term will remain in (139.3). Then under the assumption of strict unitary symmetry we obtain the asymptotic equality of the amplitudes of the elastic scattering of the π -, η -, K - and \bar{K} -mesons by baryons:

$$f_{\pi}^{\bar{}} = f_{\eta}^{\bar{}} = f_K^{\bar{}} = f_{\bar{K}}^{\bar{}}. \quad (140.12)$$

In the case of violated symmetry it is necessary to make use of formula (140.11); hence follows the relation between the amplitudes similar to the mass relation:

$$f_{\pi}^{\bar{}} + 3f_{\eta}^{\bar{}} = 2(f_K^{\bar{}} + f_{\bar{K}}^{\bar{}}). \quad (140.13)$$

In concluding this section we note that one can by the same procedure investigate the isospin symmetry violated by the electromagnetic interaction using the Hamiltonian density (135.10). We leave it to the reader to obtain on his own, as a useful exercise, the following mass formulae for the isotriplet Σ and isoquartet Δ :

$$m_{\Sigma^0} = \frac{1}{2}(m_{\Sigma^+} + m_{\Sigma^-}) \quad (140.14)$$

and

$$m_{\Delta^{++}} - m_{\Delta^-} = 3(m_{\Delta^+} - m_{\Delta^0}). \quad (140.15)$$

§141. Composite models in the unitary symmetry scheme. Quarks

At the beginning of §134 we pointed out that the π -meson (and generally speaking, also all other non-strange particles) can be conceived of as a particle

consisting of a nucleon and an antinucleon. It is natural to try to formulate an analogous 'minimum' model in which all hadrons are constructed from a small number of some particles said, in a certain sense, to be fundamental. For this it would be necessary to add to the nucleon, which is a carrier of a baryonic number and an isospin (and this means also an electric charge), at least one particle which possesses strangeness. The most economic model of such a type was proposed in 1956 by Sakata, who chose as fundamental particles p, n, Λ and $\bar{p}, \bar{n}, \bar{\Lambda}$ and assumed that there is an attraction between any fundamental baryon and antibaryon, and a repulsion between two baryons or antibaryons. The wave functions of the hadrons known at that time were constructed as follows:

$$\begin{aligned}
 \pi^+ &= p\bar{n}, & \pi^- &= \bar{p}n, & \pi^0 &= \frac{1}{\sqrt{2}}(p\bar{p} - n\bar{n}); \\
 K^+ &= p\bar{\Lambda}, & K^0 &= n\bar{\Lambda}, & K^- &= \bar{p}\Lambda, & \bar{K}^0 &= \bar{n}\Lambda; \\
 \Sigma^+ &= p\bar{n}\Lambda = \pi^+\Lambda, & \Sigma^- &= \bar{p}n\Lambda = \pi^-\Lambda, & \Sigma^0 &= \frac{1}{\sqrt{2}}(p\bar{p} - n\bar{n})\Lambda = \pi^0\Lambda; \\
 \Xi^- &= \bar{p}\Lambda\Lambda = K^-\Lambda, & \Xi^0 &= \bar{n}\Lambda\Lambda = \bar{K}^0\Lambda.
 \end{aligned} \tag{141.1}$$

This model was developed by Markov, Okun and others, and made it possible to obtain a large number of interesting physical results.

It turns out that the Sakata model fits the scheme of unitary symmetry very well, if the mass difference between the nucleon and Λ -hyperon is neglected. For this it is sufficient to assume that these particles form the triplet (1,0), and the corresponding antiparticles the conjugate triplet (0,1):

$$S_i = \begin{pmatrix} p \\ n \\ \Lambda \end{pmatrix}, \quad \bar{S}^i = (\bar{p}, \bar{n}, \bar{\Lambda}). \tag{141.2}$$

The isospin component operator (138.1) for p, n and Λ gives the correct values $+\frac{1}{2}$, $-\frac{1}{2}$ and 0, but the hypercharge operator (138.2) must be modified in such a way that instead of leading to fractional values it leads to $Y = +1, +1, 0$, respectively. We assume by definition

$$\hat{Y} = \frac{1}{\sqrt{3}}\Lambda_8 + \frac{2}{3}\hat{B}I, \tag{141.3}$$

so that instead of (138.6) we shall now have

$$Y = \frac{1}{3}(p-q) - p_3 + q_3 + \frac{2}{3}B. \tag{141.4}$$

On the basis of the requirement that the hypercharge be an integer, and taking into account that for mesons $B=0$, we arrive at the old relation $p-q=3n$, i.e. in the Sakata model these particles must also fill unitary singlets, octuplets and so on. But for baryons $B=1$, and instead of (138.7) we obtain

$$p-q=3n-2 \quad (141.5)$$

i.e. baryons must fill unitary triplets $(1,0)=3$, sextets $(0,2)=\bar{6}$, 15-plets $(2,1)=15$ and so on. Making use of formulae (138.5) and (141.4) we immediately find the values of isospin and hypercharge for the components of these multiplets (see table 7).

Table 7

Components	Y	T	Components	Y	T
φ_a	1	$\frac{1}{2}$	φ_{ab}^3	2	1
φ_3	0	0	φ_{3a}^3	1	$\frac{1}{2}$
φ^{33}	2	0	$\varphi_{ab}^6 + \frac{1}{3}[\delta_a^c \varphi_{3b}^3 + \delta_b^c \varphi_{3a}^3]$	1	$\frac{3}{2}$
φ^{3a}	1	$\frac{1}{2}$	φ_{33}^3	0	0
φ^{ab}	0	1	$\varphi_{3a}^b + \frac{1}{2}\delta_a^b \varphi_{33}^3$	0	1
$(a,b=1,2)$			φ_{33}^a	-1	$\frac{1}{2}$

The distribution of hadrons over the multiplets mentioned corresponds to their wave functions (141.1). Indeed, mesons are made up of a 'sakaton' S_i and the 'antisakaton' \bar{S}^i , and their wave functions transform according to the direct product $3 \otimes \bar{3}$ which in its decomposition (137.4a) contains just the representations 1 and 8. Baryons are made up of two sakatons and an anti-sakaton, and the decomposition (137.4c) of the direct product $3 \otimes 3 \otimes \bar{3}$ includes just the necessary representations 3, $\bar{6}$ and 15.

It is seen from the table that the Σ -hyperon must be placed at least in a sextet which includes a nucleon-like particle and a particle with $Y=+2$, $T=0$ which must have spin $\frac{1}{2}^+$. These particles have up to now not been discovered, although there are no prohibitions imposed upon their existence. The Ξ -hyperon must be contained in a 15-plet in which a large number of unoccupied places remains, and the Ω -hyperon cannot be included in any of the lower multiplets. Thus the classification of hadrons based on the Sakata model is much less satisfactory than in the eightfold way formalism.

Furthermore, it leads to a number of conclusions contradicting experiment: for example, in the Sakata model the observed process

$$\bar{p} + p \rightarrow K_L^0 + K_S^0$$

is forbidden.

Wishing to preserve all the advantages of composite models on the one hand, and those of the eightfold way formalism on the other hand, Gell-Mann and independently Zweig in 1964 proposed to renounce the modification of the hypercharge operator (138.2) and at the same time to assume that there is a unitary triplet

$$q_i = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}, \quad \bar{q}^i = (\bar{q}_1, \bar{q}_2, \bar{q}_3) \quad (141.6)$$

of particles possessing very unusual properties. From (138.4)–(138.6) and (138.9) it follows that they have the quantum numbers given in table 8 (the baryon number is by definition equal to $\frac{1}{3}$) (all quantum numbers of the anti-particles, except for T , have the opposite sign), i.e. the electric charge, baryonic number and hypercharge of these particles are fractional numbers. Their name, quarks (something incomprehensible and mystical from one of the novels of the Irish writer J. Joyce), is due to just these properties.

Table 8

Particle	Q	T	T_3	S	B	Y
q_1	$+\frac{2}{3}$	$\frac{1}{2}$	$+\frac{1}{2}$	0	$\frac{1}{3}$	$+\frac{1}{3}$
q_2	$-\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{1}{3}$	$+\frac{1}{3}$
q_3	$-\frac{1}{3}$	0	0	-1	$\frac{1}{3}$	$-\frac{2}{3}$

Mesons are made up of a quark and an antiquark, and from (137.4a) are distributed over unitary singlets and octets. If the pair $q\bar{q}$ is in a 1S_0 -state, then its total spin is equal to zero and its parity is odd, and we obtain pseudoscalar mesons with the wave functions

$$\begin{aligned}
\pi^+ &= q_1 \bar{q}_2, & \pi^- &= \bar{q}_1 q_2, & \pi^0 &= \frac{1}{\sqrt{2}} (q_1 \bar{q}_1 - q_2 \bar{q}_2); \\
K^+ &= q_1 \bar{q}_3, & K^0 &= q_2 \bar{q}_3, & K^- &= \bar{q}_1 q_3, & \bar{K}^0 &= \bar{q}_2 q_3; \\
\eta &= \frac{1}{\sqrt{6}} (q_1 \bar{q}_1 + q_2 \bar{q}_2 - 2q_3 \bar{q}_3); \\
X^0 &= \frac{1}{\sqrt{3}} (q_1 \bar{q}_1 + q_2 \bar{q}_2 + q_3 \bar{q}_3)
\end{aligned} \tag{141.7}$$

where X^0 is a resonance in the $\pi\pi\eta$ system with a mass of 960 MeV, with $T = Y = 0$, and which is a unitary singlet. If the pair $q\bar{q}$ is in a 3S_1 -state, then its spin is equal to 1⁻, and we arrive at the vector mesons ρ , K^* , ω' and φ' whose wave functions are constructed analogously to (141.7). Since for the quark $B = \frac{1}{3}$, a system of three such particles will have a baryonic number equal to 1 and, hence it is natural to identify it with a baryon (we recall that in the Sakata model baryons are made up of two particles and an antiparticle). The wave function of the system qqq transforms according to the representation $3 \otimes 3 \otimes 3$ and therefore it follows from the decomposition (137.4b) that in the quark model, as well as in the eightfold way formalism, baryons fill unitary singlets, octets and decuplets. If the spins of two quarks are parallel, then there are $9 \frac{1}{2}^+$ states, of which one is a unitary singlet, while the 8 remaining states belong to a unitary octet. If the spins of all three quarks are parallel, then we obtain $10 \frac{3}{2}^+$ states forming a decuplet.

We shall not consider the dynamic consequences of the quark model and some of its inherent difficulties, but refer the reader to the corresponding literature*. We shall dwell only briefly on the problem of the reality of the existence of these unusual particles. If quarks indeed exist, then their world must be almost independent of the ordinary world: from the fact that their charge is fractional it follows that the lightest of the quarks must be absolutely stable; they can be produced only in the form of quark-antiquark pairs, for example as a result of the bombardment of ordinary matter by cosmic rays. Therefore in the course of time the total number of quarks contained in the Earth's crust and in the waters of the oceans must increase progressively. However, numerous attempts to find these 'relic' quarks by means of precision apparatus have not yet been successful. Nor has anyone succeeded in discovering them in experiments with accelerators; only a lower

* See, for example, the reviews of E.M. Levin and L.L. Frankfurt in *Usp. Fiz. Nauk* 94 (1968) 243, and a rather popular review article of Ya.B. Zeldovich in *Soviet Phys. Usp.* 8 (1965) 489.

limit for their mass has been established: $m_q > 5 \text{ GeV}$ ($5 \times 10^3 \text{ MeV}$). This points to the fact that if hadrons are actually made up of quarks, then their binding energy must be colossal. In this situation an ever increasing number of physicists (including Gell-Mann himself) are beginning to be inclined to the idea that even if quarks do exist they cannot be in a free state, but are similar to quasi-particles, for example to phonons, in a solid. Some outstanding scientists (Heisenberg, Chew and others) disapprove of the hypothesis of quarks.

But in spite of everything the quark model is very attractive and even at the worst it is a very convenient mathematical tool for the formulation of unitary symmetry. The near future should give answer to one of the cardinal questions of the contemporary physics of elementary particles: are quarks real, and if so, in what sense, or are they a purely mathematical fiction?

§142. General appraisal of unitary symmetry

From the contents of the preceding sections it is seen that, owing to the hypothesis of the approximate invariance of the strong interaction with respect to the group $SU(3)$, the physics of elementary particles has recently made much progress. The successes of unitary symmetry are numerous and impressive.

1. All stable hadrons and low-lying resonances are distributed over unitary multiplets whose members have one and the same spin, parity and baryonic number: octet of 0^- mesons, octet of $\frac{1}{2}^+$ baryons, decuplet $\frac{3}{2}^+$, nonets of 1^- and 2^+ meson resonances and others. No particle which in principle could not be placed in one of the unitary multiplets of not too high dimension has been discovered.

2. The quark model, which makes it possible to construct all hadrons from three fundamental particles and their antiparticles, is very attractive.

3. Different mass formulae have been obtained for isomultiplets within one unitary multiplet, which are in very good agreement with empirical data.

4. A number of relations between the coupling constants of baryons with mesons have been established, which for the most part need experimental verification.

5. The relations between the cross sections for different processes have been derived; when some subsidiary facts are accurately taken into account, none of them is in sharp disagreement with experiment.

6. Taking into account the electromagnetic interaction, mass formulae

have been obtained for individual members of the isomultiplets constituting a unitary multiplet. For example, on the basis of the theoretical relation

$$(m_n - m_p) - (m_{\Xi^0} - m_{\Xi^-}) = m_{\Sigma^-} - m_{\Sigma^+} \quad (142.1)$$

the sign of the mass difference between the Ξ^0 and Ξ^- -hyperons was predicted. This prediction was subsequently confirmed experimentally.

7. The relations between the magnetic moments of the baryons belonging to the same unitary multiplet were derived. In particular, the following equalities were shown to hold for the members of the octuplet $\frac{1}{2}^+$:

$$\mu_p = \mu_{\Sigma^+}, \quad \mu_{\Xi^-} = \mu_{\Sigma^-}, \quad \mu_n = 2\mu_{\Lambda} = -2\mu_{\Sigma^0}. \quad (142.2)$$

8. Finally, Cabbibo succeeded in including the weak interaction also in the scheme of unitary symmetry, by which this theory acquired a certain harmony and completeness.

The first five points were discussed in §138–141. We shall not dwell on the remaining points, but refer the reader to the literature*.

On the other hand, the hypothesis of unitary symmetry has a number of essential shortcomings.

1. First of all, the group theoretical scheme does not contain any elements of dynamics, and it must only be a part of a theory of elementary particles which is still to come.

2. The question of the nature of unitary symmetry and its violation remains open. In connection with this different points of view have been put forward:

(a) Unitary symmetry is a fundamental property of the strong interaction, and is inherent to it in the same way as, say, its invariance with respect to transformations of the Lorentz group. In this case it is violated either by a moderately strong interaction whose coupling constant is of the order of 0.1 (we recall that the coupling constant of the strong interaction is equal to about 14), or owing to its interaction with the vacuum state which may have a complex structure (the so-called spontaneous violation of symmetry).

(b) Unitary symmetry is approximate in its very nature, 'because of a complex concurrence of different factors'**. In this case, one need not raise the question of the nature of the violation of SU(3) invariance, and the require-

* See, for example, the monograph of Nguyen van Hieu, *Lektsii po teorii unitarnoi simmetrii elementarnikh chastits* (Lectures on the theory of elementary particles) (Atomizdat, Moscow, 1967).

** G. Chew, *The analytic S-matrix* (Benjamin, New York, 1966).

ment that a unitary triplet of particles (quarks) necessarily exists makes no sense.

(c) There exists a strictly unitary symmetric superstrong interaction, and SU(3) invariance is violated as a result of switching on a real strong interaction. Such a situation is analogous to that in the case of isospin symmetry.

3. The quark model in its initial version described in §141 contains a number of logical difficulties. One of these lies in the fact that in trying to construct hadrons from three quarks the spatial-spin part of the wave function turns out to be antisymmetric with respect to permutation of the quarks, which is unusual for the lowest state. In trying to overcome this difficulty the quark was assigned a new quantum number, which is equivalent to the consideration of not three but nine quarks, whence the harmony and elegance of the model are lost. There were even propositions to assume that quarks do not obey Fermi–Dirac statistics but a certain new type (so-called parastatistics) in which occupation numbers may take on, say, the values 0, 1 and 2.

4. The unitary symmetry scheme has a number of other shortcomings, which are not so fundamental but are still important. We shall only enumerate them:

(a) The situation concerning the distribution of higher resonance states over multiplets is not quite clear. For example, the problem as to where the particle Λ_{1405} , having spin $\frac{1}{2}^-$, is to be placed has not been elucidated.

(b) The problem of ω – φ mixing has not been finally solved.

(c) The question remains open as to why the mass formula for baryons involves the mass itself, while the mass formula for mesons involves the square of the mass.

(d) There is contradiction between the high accuracy of the mass formula for the decuplet $\frac{3}{2}^+$ and the poor relations for the probabilities of decay of these resonances, and so on.

It should also be noted that for a number of reasons the framework of the group SU(3) turns out to be too narrow:

1. It does not involve the baryonic number.

2. The parameters of ω – φ mixing are not predicted but are introduced into the theory from outside.

3. There are a number of intermultiplet relations pointing to the correlation of unitary quantum numbers with ordinary spin. Thus, for example, the parameters contained in the Okubo formula (140.10) for the octet and decuplet of baryons turn out to be almost equal; the following mass formula is valid

$$m_K^2 - m_\rho^2 = m_K^2 - m_\pi^2 \quad (142.3)$$

and so on.

4. Some physicists are not content with the fact that the charge of quarks is a fractional number, which calls for a correction of the Gell-Mann—Nishijima relation by introducing into it a new quantum number, and this means increasing the rank of the basic symmetry group.

In trying to partially overcome the shortcomings mentioned a theoretical scheme was formulated invariant with respect to the group SU(6), which describes at the same time the unitary symmetry and spin properties of particles. Pseudoscalar and vector mesons fill the 35-dimensional multiplet of this group (every 0^- meson has one spin state, while the 1^- meson has three spin states, so that if in addition a unitary scalar 1^- particle is introduced, then we shall obtain in all $8 \times 1 + 8 \times 3 + 1 \times 3 = 35$ members of the multiplet), while $\frac{1}{2}^+$ and $\frac{3}{2}^+$ baryons fill the 56-plet ($8 \times 2 + 10 \times 4 = 56$). The most impressive result of the group SU(6) is the formula for the ratio of the magnetic moments of the proton and neutron:

$$\mu_n/\mu_p = -\frac{2}{3} \quad (142.4)$$

(the experimental value being -0.68). But the group SU(6) is essentially non-relativistic and is unsuitable, for example, for a description of the processes of scattering of particles. However, in the attempt to make it relativistic insuperable difficulties have arisen associated with the probabilistic interpretation of the corresponding quantum-mechanical scheme.

The trend of further development of the theory is to be indicated by experiment.

SUBJECT INDEX

- Absorption, 387
 - coefficient, 389
 - cross section, 389
 - of light, 442
 - of neutrons by nuclei, 393
- Adiabatic approximation, 229
 - operator, non-, 231
- Adjoint operator, 65
- Airy function, 54
- Amplitude of a field, 439
- Angular momentum, addition of, 204
 - — conservation law, total, 112, 249
 - —, motion of particle with, 130
 - — of atom, 293
 - — of N -particle system, 250
 - — operator, 101, 199
 - — —, matrix elements of, 200
 - — projection, 202
 - —, total, 248
- Annihilation operator, 198, 425
- Anticommuting operators, 62
- Antiparticle, 497
- Antisymmetric wave function, 256
- Associative law, 170
- Atom, electron density in, 289
 - in magnetic field, 308
- , ordering of energy levels of, 297
- , radius of, 289
- , statistical model of, 285
- Atomic form factor, 355
 - shell, 297
- Austausch integral, 267
- Band, 343
- Barrier, rectangular one-dimensional, 45
 - , scattering by, 367
- Baryon, 550
 - meson interaction, 596
 - number, 551
 - resonance, 550
- β -decay, 506
- Bilinear transformation, 246
- Bohr magneton, 236
 - quantization rule, 156
 - radius, 141
- Born approximation, 350
 - —, applicability of, 352
- Bose—Einstein particles, 257, 424
- Bosons, 257, 424
 - , wave function for, 257
- Boundary conditions on wave function, 29
- Bound states and poles in scattering matrix, 413
- Box normalization, 89
- Bra vector, 188
- Breit—Wigner formulae, 399
- Bremsstrahlung, 533
- Broglie wave, de, 15
 - 's formula, de, 15
- Canonical transformation, 177
- Capture, neutron, 400
- Causality, law of, 25
- Centrally symmetric field, 122
 - — —, conservation laws in, 122
 - — —, quasi-classical method in, 162
- Centre-of-mass system, 347
- Centrifugal energy, 125
- Channel, open, 385
 - , reaction, 385, 404
- Charge conjugation operator, 490
 - density and Dirac equation, 482
 - — and Klein—Gordon—Fock equation, 472

- Chebyshev-Hermite polynomials, 40
- Chemical binding, 327
- Chronological operator, 410
- Classical limit, 145
 - - of cross section, 382
 - - of scattering, 378
 - - of scattering amplitude, 380
- Clebsch-Gordon coefficients, 205
- Coherent scattering, 456
- Collision broadening, 466
 - of identical particles, 254
 - of hard identical balls, 369
 - of spin $\frac{1}{2}$ particles, 370
 - of the second kind, 346
 - vector, 351
- Commutation relations of angular momentum, 102
- Commutator, 62, 78, 85
- Commuting operators, 61, 78
- Complete set of eigenfunctions, 70
 - - of quantum-mechanical quantities, 80
- Composite models, 605
- Compton scattering, 531, 535
- Conjugate matrix, 172
- Conservation law, 111
 - - and interactions, 556
 - - and light emission, 445
 - - and selection rules, 452
 - -, particle, 32
 - - of symmetry property, 257
- Constant of the motion, 111
- Continuous energy spectrum, 54
 - matrix, 183
- Coordinate representation, 75
 - operator, 88, 185
 - -, eigenvalues of, 91
- Corpuscles, 4
- Correspondence principle, 22, 72
- Coulomb field, 135
 - system, 136
- Coupling constant, 506
 - , jj -, 298
 - , Russel-Saunders, 298
- CPT* invariance, 509
- Creation operator, 198, 425
- Cross section and partial wave theory, 361
 - - and scattering matrix, 404
 - - and unitary invariance, 597
 - -, classical limit of, 382
 - -, convergence of partial, 362
 - -, differential, 345
 - - for photoelectric effect, 456
 - - for Raman scattering, 461
 - - for scattering of identical particles, 368
 - -, inelastic scattering, 388
 - -, partial, 361
 - -, - inelastic scattering, 387
- Current density and Dirac equation, 482
 - - and Klein-Gordon-Fock equation, 472
 - -, probability, 31
 - - vector for particles with spin, 252
- D-coupling, 596
- Degeneracy, 37, 65, 214
 - , accidental, 140
 - in one-dimensional motion, 53
- Degenerate level, perturbation theory of, 214
- Detailed balance, 421
- Determinant, 171
- Deuteron, 314, 577
 - , model for, 317
 - , potential energy of, 315
 - , quadrupole moment of, 317
 - , states of, 316
- Diagonalization, 181
- Diagonal matrix, 167
- Diamagnetic susceptibility, 314
- Diatomic molecule, 323
 - -, rotation of, 339
 - -, vibration of, 340
- Dielectric constant, 461
- Differential cross-section, 345
 - operator, 60
- Dimension of representation, 558, 565
- Dipole-dipole interaction, 334
 - moment of atom, electric, 298
 - radiation, 447
 - -, magnetic, 449
 - transition, 447
 - -, selection rules for, 451
- Dirac bispinor, 481

- δ -function, 69
- equation, 478, 518
- - and electromagnetic field, 523
- - and free particle, 483
- - and hydrogen atom, 497
- - and Lorentz transformation, 501
- - and Pauli equation, 494
- - and rotation, 499
- - and spin particle, 492
- -, non-relativistic limit of, 496
- notation, 188
- perturbation theory, 220
- Discrete energy spectrum, 53
- Dispersion formula, 459
- relations, 415
- Displacement operator, 90
- Doppler broadening, 466
- Double scattering, 376
- Duality, wave-particle, 5
- Dyson's formula, 411

- Ehrenfest theorem, 110
- Eigenfunction, 64, 180
- in centrally symmetric field, 123, 126
- Eigenvalue, 64, 180
- Eightfold way, 591
- Elastic scattering, 346
- Electric dipole moment of atom, 298
- Electromagnetic field, Hamiltonian of, 436
- -, momentum of, 438
- - with electron, 440
- interaction, 554
- vacuum, 543
- -, electron in, 543
- Electron, classical radius of, 455
- , equivalent, 295
- states in atom, 293
- Elementary particles, properties of, 550
- -, stable, 552
- η -meson, 550
- Emission of light, 442
- , spontaneous, 445
- , stimulated, 444
- Energy conservation, 112
- levels of the atom, ordering in, 290
- - of helium atom, 278
- spectrum in centrally symmetric field, 128
- - of hydrogen atom, 139
- Euler angles, 244
- Exchange energy and chemical binding, 332
- - in nuclei, 273
- -, saturation of, 271
- integral, 267
- interaction, 264, 269
- operator, particle, 256
- Excited state, 33
- External field, uniform, 54
- line, 530

- F-coupling, 596
- Fermi constant, 555
- Dirac particles, 257, 430
- Fermions, 257, 430
- , wave function for, 257
- , wave function for two, 261
- Feynman diagram, 529
- formulation and perturbation theory, 232
- Filled shells and Thomas-Fermi model, 289
- Final state, 401
- Forbidden transition, 117, 448
- Form factor, atomic, 355
- - of hydrogen atom, 356
- Four-fermion interaction, 513
- Franck-Hertz experiment, 4
- Free particle and Dirac equation, 483
- -, system of, 13, 56

- Gell-Mann mass formula, 603
- Nishijima relation, 554
- Generators, group, 559
- , representation, 560
- γ -matrices, 482, 502
- Gravitational interaction, 556
- Green's function, 62
- - and Dirac equation, 518
- - for two-particle system, 524
- -, relativistic, 526
- Ground state, 33
- Group generators, 559
- representation, 558
- velocity, 15

- Hadenon, 550
- Hadron, 550
 - , isomultiplets of, 568
- Hamiltonian for electromagnetic field, 436
 - in second quantized representation, 428
 - operator, 92
- Harmonic oscillator, linear, 38, 195
 - —, three-dimensional, 43
- Hartree equations, 284
 - Fock equations, 285
 - — method, 282
- Heisenberg representation, 192
 - 's uncertainty relations, 20, 82
- Helicity operator, 512
- Helium atom, 276
 - —, eigenstates of, 278
 - —, energy levels of, 278
- Hermite polynomials, Chebyshev, 40
- Hermitian matrix, 167, 173
 - operator, 65
 - —, eigenvalues of, 67
- Hilbert space, 175
- Homopolar molecules, 327
- Hybridization, 334
- Hydrogen atom, 121, 135, 217
 - —, form factor of, 356
 - —, relativistic corrections, 498
 - — with finite nucleus, 217
 - molecule, 327
 - spectrum, 141
- Hypercharge, 554, 585
- Identical balls, collision of hard, 369
 - particles, collision of, 254
 - —, scattering of, 368
 - —, wave function for, 255
 - —, wave function for two, 261
- Impact parameter, 363
- Inelastic scattering, 346, 385
 - — cross section, 388
- Initial state, 401
- Integral operator, 60
- Interaction constant, 440
 - , dipole-dipole, 334
 - , electromagnetic, 554
 - , exchange, 268
 - , four-fermion, 513
 - , gravitational, 556
 - , interatomic, 334
 - of atoms in different states, 335
 - , quadrupole-quadrupole, 335
 - , relativistic, 526
 - representation, 194
 - , spin-orbit, 377
 - , strong, 476, 555
 - , van der Waals, 335
 - , weak, 504, 555
- Intermediate state, 228
- Internal line, 530
- Invariant amplitudes, method of, 583
 - operators, 561
- Inverse matrix, 171
 - operator, 62
- Inversion operator, 114, 325
 - — and spinor transformation, 247
- Inverted multiplets, 295
- Ions and the Thomas-Fermi model, 291
- Irreducible representation, 558
- Isogroup, 562
- Isometric nuclei, 452
- Isomultiplets, 551, 567
 - of hadrons, 568
- Isospin, 551, 568
 - component operators, 567
 - conservation, 577
 - , third component of, 568
- jj*-coupling, 298
- Kernel of integral operator, 60
- Ket vector, 188
- Ξ -hyperon, 550
- Klein-Gordon-Fock equation, 468
 - — — — and charge density, 472
 - — — — and current density, 472
- Klein-Nishina formula, 542
- K-meson (kaon), 507, 550
- Kronecker symbol, 68
- Laboratory system, 347
- Laguerre polynomial, generalized, 139
- Lamb shift, 543
- Landé *g*-factor, 311
- Laplacian, 60

- Λ -hyperon, 550
- Lepton, 515, 549
 - number, 551
- Lie group, 559
 - —, dimension of, 559
- Light absorption, 442
 - emission, 442
 - scattering by atoms, 456
- Linear operator, 60
- Linewidth, natural, 462
- Lorentz transformation and Dirac equation, 501
- Magic numbers, 322
- Magnetic field, atom in, 308
 - —, particle in, 93
 - moment operator, intrinsic, 251
 - quantum number, 104
- Many-particle system, 55, 94
 - — —, non-interacting, 13, 56
 - spin particle system, 253
- Matrices, product of, 169
 - , sum of, 169
- Matrix, 164
 - calculus, 168
 - element, 165
- Maxwell's equations, 432
- Mean value, 76
- Measurement, 23, 74, 119
 - , simultaneous, 79
- Mendeleyev periodic system, 299
- Mesohydrogen, 142
- Meson, 550
 - baryon interaction, 596
 - resonance, 550
- Metal, electrons in a, 230
- Metastable state, 452
- Microparticles, 4
- Mixed state, 81
- μ -meson (muon), 505, 549
- Molecular and atomic states, 337
 - spectrum, 342
 - states, 325
- Molecule, diatomic, 323
 - , geometric form of, 333
- Momentum conservation, 112
 - operator, 73, 87
 - — and infinitesimal translation, 90
 - —, eigenfunctions of, 88
 - representation, 88, 185
 - transfer vector, 351
- Moseley law, 305
- Multiplet, 558
 - , inverted, 295
 - , normal, 295
- Multiplicity, 561
 - of degeneracy, 37
- Muon, 505, 549
- Neutrino, 549
 - , two-component, 509
- Neutron absorption by nuclei, 393
 - capture, 400
 - nucleus scattering, 377, 391
 - — —, inelastic, 397
 - proton scattering, 581
- Normal multiplets, 295
- Normalized eigenfunction, 68
- Nuclear field, 473
 - forces, theory of, 272
 - shell, 320
- Nuclei, self-consistent field in, 320
- Nucleon-antinucleon system, 573
 - charge, 569
 - nucleon scattering, 580
 - — system, 574
 - π -meson interaction, 578
 - — scattering, 580
 - — system, 575
 - resonances, 572
- Null matrix, 168
- Number operator, 425
- Occupation number representation, 423
- Ω -hyperon, 550, 594
- Okubo mass formula, 603
- One-dimensional motion, 52
- Operator, 60, 188
 - and physical variables, 71
 - in second quantization-representation, 425
 - product, 61
- Optical model for scattering, 391
 - theorem, 390
- Orbital angular momentum quantum number, 105

- Orthogonal eigenfunctions, 67
- Orthohelium, 277
- Orthostates, 277
- Oscillator strength, 461

- Pair production frequency, 546
- Parahelium, 277
- Parastates, 277
- Parity, 115, 551
 - conservation, 117
 - non-conservation, 507
- Partial cross section, 361
 - -, convergence of, 362
 - inelastic cross section, 387
 - wave theory, 357
 - width, 400
- Particle in field, 28
 - , stable, 549
- Paschen-Back effect, 313
- Pauli equation, 251
 - - and Dirac equation, 494
 - - for many-particle system, 253
- Pauli-Lüders theorem, 551
- Pauli operators, 240
 - principle, 260
- P-branch, 342
- Penetration, 53, 158
- Periodic boundary conditions, 89
 - system, 299
- Perturbations, adiabatic theory of, 229
 - calculation for helium atom, 278
 - corrections to energy, 212, 213
 - - to wave function, 212, 213
 - operator, 210
 - theory and scattering matrix, 408
 - - and Feynman formulation, 232
 - -, Dirac, 220
 - - of degenerate level, 214
 - , time-dependent, 219
 - , transition due to, 222
- Phase shift, 358
 - -, uncertainty in, 407
 - velocity, 15
- Photoelectric effect, 452
 - -, cross section for, 456
- Photon, 432, 549
 - , interaction of, 435
 - spin, 435, 442
 - transition, one-, 442
 - , wave function of, 433
- Physical variables and operators, 71
- π -meson (pion), 274, 473, 550, 571
 - nucleon interaction, 578
 - - scattering, 580
 - - system, 575
 - production, 582
- Poisson bracket, 84
- Polarization and scattering, 371
 - of scattered beam, 376
 - vector, 372
- Polar spinor, 248
- Poles, false, 413
 - in scattering matrix and bound states, 413
- Positron, 487
 - , Feynman diagram for, 531
 - state, 523
- Potential, complex, 390
 - well, 32, 155
- Principal quantum number, 138
- Probability density, 11
- Propagation function of the virtual photon, 507
- Proton-neutron scattering, 581
 - π -meson scattering, 582
- Pseudo-spinor, 248
- Pure state, 80

- Q-branch, 343
- Quadrupole radiation, 449
 - quadrupole interaction, 335
- Quantum-mechanical system, properties of, 35
- Quark, 607
- Quasi-classical approximation, 145
 - -, validity of, 150
 - wave function, 153

- Radial quantum number, 138
 - wave equation, 125
- Raman effect, 459
 - scattering, 456
- Rank, group, 561
 - of a matrix, 171
- R-branch, 342
- Reaction channel, 385, 404

- Reciprocity theorem, 420
- Rectangular potential well, one-dimensional, 32
 - - -, three-dimensional, 36
- Reducible representation, 558
- Reflection coefficient, 47
- Refractive index, 461
- Relativistic interaction, 526
- Renormalization, 535
- Representation, coordinate, 75
 - , dimension of, 558
 - , general, 75
 - , generators, 560
 - , irreducible, 558
 - of a group, 558
 - , operator, 177
 - , scalar, 563
 - , spinor, 563
 - , vector, 566
- Resonance, 549, 553
 - , baryon, 550
 - , fluorescence, 462
 - , meson, 550
 - , scattering, 364, 399
 - , transition, 226
- Reversal in time, 30
- Rotation matrix, 242
 - of diatomic molecules, 339
 - operator, 107
- Russel-Saunders coupling, 298
- Rutherford formula, 355
- Rydberg constant, 141

- Sakata model, 605
- Saturation of exchange energy, 271
- Scalar representation, 563
- Scatterer, 345
- Scattering amplitude, 347
 - - in Born approximation, 350
 - - in classical limit, 380
 - - and partial wave theory, 360
 - and polarization, 371
 - by barrier, 367
 - by spherical well, 364
 - , classical limit of, 378
 - , coherent, 456
 - cross section, differential, 345
 - - -, inelastic, 388
 - - -, partial inelastic, 387
 - , double, 376
 - , elastic, 346
 - , inelastic, 346, 385
 - matrix, 401
 - , analytic properties of, 412
 - - and perturbation theory, 408
 - - and time reversal, 418
 - - and transition probability, 403
 - , neutron-nucleus, 377, 391
 - , neutron-proton, 581
 - of charged particles by atoms, 354
 - of identical particles, 368
 - of spin $\frac{1}{2}$ particles, 370
 - of slow particles, 394
 - , π -meson-proton, 582
 - process, 345
 - , Raman, 456
 - resonance, 364, 399
 - theory, partial wave, 357
- Schrödinger equation, 29
 - , integral form of, 96
 - - in matrix form, 186
 - representation, 191
- Screening of nuclear charge, 356
- Second quantization, 423
- Secular equation, 216
- Selection rules, 342, 449, 561
- Self-consistent field in nuclei, 320
 - - method, 282
- Shell, atomic, 297, 299
 - , closed, 297, 321
 - , nuclear, 320
- Shur lemma, 561
- Σ -hyperon, 550
- Sign function, 410
- Spectral linewidth, 118
 - term, 140
- Spectrum of an operator, 64
 - , molecular, 342
- Spherical well, 132
 - , scattering by, 364
- Spin, classical limit of, 240
 - component operators, 239
 - - -, eigenvalues of, 241
 - - -, rotation of, 242
- , electron, 234
- operator, 236

- orbit interaction, 294, 377
- particle and Dirac equation, 492
- - system, many-, 253
- quantum number, 235
- , wave function for particle with, 237
- Spinor, 245
- , polar, 248
- , pseudo-, 248
- representation, 563
- Spontaneous emission, 445
- Spread of wave packet, 19
- Standard conditions, 64
- Stationary states, 28, 94
- Statistical model of the atom, 285
- Stern-Gerlach experiment, 5, 235
- Stimulated emission, 444
- Strangeness, 554, 585
- Strange particle, 554
- Strong interaction, 476, 555
- Sub-levels, 214
- SU(2) group, 562
- -, representations of, 563
- SU(3) group, 586
- -, invariant operators of, 589
- - representations, 587
- Superposition principle, 18
- Symmetric wave function, 256
- Symmetry group, 557
- transformations, 557
- Thomas-Fermi equation, 287
- - model and filled shell, 289
- - - and ions, 291
- - - and quasi-classical approximation, 292
- Thomson formula, 543
- Threshold approximation, 394
- Time-energy uncertainty relation, 117
- Time-ordering operator, 410
- Time-reversal, 418
- Trace of a matrix, 171
- Transition amplitude, 96, 528
- -, Fourier transform of, 99
- -, integral equation for, 101
- due to perturbations, 222
- due to time-independent perturbations, 227
- line, shape of, 466
- probability, 222, 226
- - and scattering matrix, 403
- via intermediate states, 228
- Transmission coefficient, 47, 51, 160
- Transpose of an operator, 66
- Trial function, 281
- Triangle relation, 583
- Tunnel effect, 48, 51
- Turning point, 148
- -, wave function in, 150
- Two-particle system, 57
- Uncertainty relations, 19, 82
- -, time-energy, 117
- Unit matrix, 168
- Unitary group, 586
- invariance and cross sections, 597
- matrix, 173
- symmetry, 609
- -, violation of, 601
- transformation, 177
- Vacuum, 438, 543
- , electron in, 544
- Valence, 306, 332
- Variational principle, 280
- - and Hartree-Fock approximation, 282
- Variation of constants method, 220
- Vector, n -dimensional, 174
- representation, 566
- Vibration of diatomic molecules, 340
- Virtual photon, propagation function of, 527
- state, 228
- Vortex, 529
- Waals interaction, van der, 335
- Wave function, 10
- -, boundary conditions on, 29
- - for bosons, 257
- - for fermions, 257
- - for identical particles, 255
- - for particle with spin, 237
- - for two fermions, 261
- -, N -particle, 12
- packet, 19
- Weak interaction, 504, 555

Weight diagram, 594

Wentzel-Kramers-Brillouin method,
145

Wigner formula, 366

Yukawa potential, 475

Zeeman effect, 307

Zero-point energy, 43

— -- oscillations of electromagnetic
field, 438, 544

© NORTH-HOLLAND PUBLISHING COMPANY, 1973

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior permission of the Copyright owner.

Library of Congress Catalog Card Number: 68 54501

ISBN North-Holland, complete set: 0 7204 0176 3

Vol. 4: 0 7204 0202 6

Printed in The Netherlands

Title of the Russian edition:

KURS TEORETICHESKOJ FIZIKI

Russian edition published by:

IZDATELSTVO 'NAUKA', GLAVNAJA REDAKCIJA,

FIZIKO-MATEMATIČESKOJ LITERATURY (MOSKVA, 1971)

Publishers:

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM

Sole Distributors for the Western Hemisphere:

WILEY INTERSCIENCE DIVISION

JOHN WILEY & SONS, INC. – NEW YORK

ISBN Wiley Interscience, Vol. 4: 0-471-53116-2

FOREWORD

The first Russian edition of 'Theoretical Physics', which appeared in 1962, has been widely used as a textbook.

Numerous comments from colleagues, lecturers and students have been taken into account in preparing this new edition, which is the first one in English and which will also appear as the second Russian edition.

The material has now been divided into 4 volumes covering the following subjects

Volume 1

- Part I Theory of the Electromagnetic Field
- Part II Theory of Relativity

Volume 2

- Part III Statistical Physics
- Part IV Electromagnetic Processes in Matter

Volume 3

- Part V Quantum Mechanics

Volume 4

- Part VI Quantum Statistics and Physical Kinetics

The rapid development of physics and the present wide interest in non-equilibrium and non-stationary processes has compelled us to expand the section on physical kinetics. It has also been transferred to the end of Volume 4 as it is practically impossible to expound this topic without using quantum mechanics.

Part IV — 'Electromagnetic Processes in Matter' — has been substantially revised. Interest in this field has increased recently, mainly in connection with the study of plasmas and plasma-like media, which now have sections devoted to them.

The methods of calculating electrostatic and direct-current fields, and other problems of classical electrodynamics in a medium, are covered very briefly as we have assumed that students will be able to consult the many monographs and handbooks on general physics, electrical- and radio-technology, and the equations of mathematical physics.

As for other modifications and additions, we should draw attention to the introduction of tensor notation, to new ideas in the theories of relativity and electromagnetic fields, the broadening of the introduction to the theory of probability, a brief presentation of the method of correlation functions in statistical physics, the exposition of the thermodynamic theory of ferromagnetism and the theory of propagation of electromagnetic waves in plasma. A number of paragraphs have been rewritten. We have tried to bring the content of the book even closer to the interests of present-day theoretical physics.

The general level of the book has been preserved and it is still intended to form an introduction to theoretical physics. Problems requiring the use of cumbersome or special mathematical apparatus are still excluded, and the most difficult sections are marked by an asterisk. These may be skipped at will, since there is no reference to them in the main text.

In conclusion we would like to express our gratitude to all those who helped us in preparing this book, in particular to A.M. Brodsky, A.M. Golovin, B.M. Grafov, R.R. Dogonadze, V.S. Krylov and especially V.S. Markin and V.V. Tolmachev. I.V. Savelyev discovered a number of misprints which have now been corrected.

L.D. Konkina helped us in editing the manuscript.

We are grateful to the readers and students who used the first Russian edition of the book for sending us their valuable comments which have been taken into account in this edition.

August 1970

FOREWORD TO THE FIRST RUSSIAN EDITION

The continuous development of theoretical physics and the regular expansion of its areas of application create increasing demand for textbooks and manuals.

The rapid development and the complexity of the most recent experimental methods of physical investigation, and the corresponding development and extension of the mathematical apparatus of theoretical physics, have meant that one man usually cannot combine the two methods of investigation. The end of the 19th century and particularly the 20th century therefore saw physicists divided into 'experimentalists' and 'theoreticians', the latter studying physical laws by means of the mathematical methods of theoretical physics.

Obviously, a background in theoretical physics is essential in the education of experimental as well as theoretical physicists.

The experimental and theoretical methods of physical investigation have penetrated into a number of branches of science related to physics (physical chemistry, biophysics, geophysics, astrophysics, and so on) and into technology (metal physics and metallurgical science, thermophysics, electrical technology, radiotechnology, computation, the instrument-making industry etc.). Workers in these branches of science and technology also need a certain minimum knowledge of theoretical physics.

The compilation of a modern textbook on theoretical physics is inevitably associated with certain logical and methodological difficulties. It is impossible at present to divide theoretical physics into classical and quantum parts so that it is also impossible to divide it into separate chapters and sections. For example, the exposition of statistical physics without taking into account the quantum properties of atomic systems is impossible, for it would mean that the general theory remained without practical application. In the theory of electromagnetic processes in matter one has of necessity to make use of the ideas of statistical physics, and so on. It may be that the maximum consistency of composition would be obtained if the book were founded on

quantum mechanics but this is completely inadmissible in a book intended as an introductory treatise. Quantum mechanics requires a certain preparedness and the student must be convinced of the necessity of renouncing obvious classical representations. Compromise solutions, which have justified themselves during many years of teaching theoretical physics at the Moscow Engineering-Physical Institute and Moscow State University, are therefore inevitable.

The following general principles have been applied.

(1) The book is written as an introduction to theoretical physics so that aspects requiring the use of cumbersome or special mathematical apparatus have not been included.

(2) As it is to be used for a systematic study of the subject the course is a unique whole and all material necessary for understanding the later sections is contained in the earlier ones.

(3) It would not be feasible to elucidate experimental facts in addition to problems concerning purely theoretical physics. However, physics is a single science, and an attempt to expound the theoretical aspects without taking experiment into account would be quite wrong. The reader is assumed to have some basic experimental knowledge from university courses in general and atomic physics so that we have confined ourselves to references and, in a few instances, to a schematic description of basic experiments.

(4) The acquaintance assumed with general courses in general and atomic physics has allowed us to rely on a certain (very restricted) knowledge of quantum mechanics in our treatment of statistical physics.

(5) Classical mechanics usually forms a separate course so that this topic has been omitted although detailed reference has been made to handbooks of mechanics.

(6) The book similarly does not cover hydrodynamics, aerodynamics, the theory of heat transfer, or problems related to electrical- and radio-technology.

(7) Detailed reference is made to mathematical manuals. The mathematical apparatus utilized, except in the sections marked by an asterisk, is covered by the usual courses in analysis. In the case of quantum mechanics, however, the mathematical apparatus has been included, since it is of a specific character and is not taught in traditional mathematical courses.

(8) As the book is intended as a systematic course in theoretical physics no attempt has been made to achieve the same level of accessibility in all sections. It is a well-known fact that a student's comprehension and assimilation of difficult material increases as a course progresses, and that this is also true for the associated mathematical apparatus. Moreover, experi-

mental physicists will constantly encounter new problems in quantum mechanics which can only be handled using advanced methods of treatment. The section on quantum mechanics (Part V) therefore deals with some topics having a more advanced character than those in other sections. The analysis of applications of the kinetic equations is similarly treated rather extensively.

The uniqueness of the book's objectives has affected the content of individual sections, so that some topics in modern physics have been included at the expense of more traditional material.

Part I contains the foundations of the theory of the electromagnetic field in a vacuum, based on the system of Maxwell-Lorentz equations. A basic knowledge of electromagnetism is assumed. The focus of attention is the theory of radiation and the motion of charged particles in external fields.

In Part II, devoted to the theory of relativity, a four-dimensional form of representation is adopted which not only corresponds to the spirit of the theory but also predominates in contemporary literature. The problems of dynamics in the theory of relativity are treated in some detail. A number of the most recent applications of the theory of relativity, particularly those related to nuclear physics, are covered here for the first time in a textbook.

Part III is a revised version of Levich's 'Introduction to Statistical Physics' and treats statistical physics and the fundamentals of statistical thermodynamics. Classical thermodynamics would require too much space, and did not seem indispensable.

Part IV contains the theory of electromagnetic processes in matter. Relatively little attention is paid to problems in theoretical electrical- and radio-technology. The phenomenological theory of electric and magnetic properties of matter is analyzed in some detail, and the notion of the physics of the plasma state of matter is given.

In Part V the basic ideas of present-day relativistic quantum mechanics are included as well as the traditional problems of non-relativistic quantum mechanics. Applications to solid-state theory are considered at length.

Part VI contains the essential concepts of physical kinetics, which are not usually presented in a general course on theoretical physics.

The experience of teaching theoretical physics shows that the greatest difficulties are often encountered not in understanding new physical ideas but in the actual mathematical treatments. All mathematical operations have therefore been performed in sufficient detail.

For convenience we have presented a brief derivation of those formulae of

vector analysis which are encountered throughout, as well as the necessary data on Fourier integrals and δ -function theory.

The numbering of formulae and sections starts afresh in each Part and references to appendices have been given Roman numerals.

The author hopes that the readers, after making themselves familiar with the foundations of theoretical physics expounded in this book, will be able to proceed to a more profound study using the many-volume treatise of Landau and Lifshitz. The scientific and educational ideas of their work were of great influence on the author, who is a disciple of Landau.

Parts I–IV and Part VI were written by B.G. Levich. Part V was written by Y.A. Vdovin and V.A. Myamlin under the general scientific guidance of B.G. Levich. Chapter XV of Part V was written by A.I. Naumov.

The author expresses his gratitude to the colleagues who read the book and the manuscripts, and made a number of valuable remarks: B.M. Grafov, R.R. Dogonadze, V.A. Kiryanov, V.S. Krylov, V.S. Markin, V.P. Smilga, Y.A. Chizmadzhev and Y.I. Yalamov.

The creation of a textbook on theoretical physics sufficiently comprehensive in content and clear in presentation is a very complex task. The author is therefore conscious of the fact that shortcomings and errors will be discovered and would be grateful to receive an account of them which can be taken into consideration in the next edition of the book.

Theoretical Physics: Outline of Vols. 1—4

Volume 1

Part I Theory of the Electromagnetic Field

- Chapter 1 General theory of the electromagnetic field
- 2 The electrostatic field
 - 3 The quasistationary magnetic field
 - 4 The electromagnetic field of arbitrarily moving charges
 - 5 Radiation theory
 - 6 Electromagnetic field in a vacuum and electromagnetic wave scattering
 - 7 The motion of particles in electromagnetic fields

Part II Theory of Relativity

- Chapter 1 General principles of the theory of relativity
- 2 Relativistic mechanics
 - 3 Relativistic electrodynamics

Appendix I, II and III

Subject index

Volume 2

Part III Statistical Physics

- Chapter 1 The basic concepts of the theory of probability
- 2 The kinetic theory of gases

- 3 Statistical distribution
- 4 Statistical and phenomenological thermodynamics
- 5 Ideal gases
- 6 Systems of interacting particles
- 7 Crystals
- 8 The theory of fluctuations
- 9 Systems with a variable number of particles
- 10 Statistical distributions in quantum statistics and some of their applications

Part IV Electromagnetic Processes in Matter

- Chapter 1 Electromagnetic fields in matter
- 2 Electrostatics
 - 3 Direct electric current and the magnetic properties of matter
 - 4 Quasistationary electromagnetic fields
 - 5 High-frequency fields
 - 6 Matter in the plasma state

Appendix IV

Subject index

Volume 3

Part V Quantum Mechanics

- Chapter 1 The basic concepts of quantum mechanics
- 2 The Schrödinger equation
 - 3 The mathematical apparatus of quantum mechanics
 - 4 Motion in a centrally symmetric field
 - 5 The quasi-classical approximation
 - 6 The matrix form of quantum mechanics
 - 7 Perturbation theory
 - 8 Spin and identity of particles
 - 9 Applications of quantum mechanisms to atomic and nuclear systems
 - 10 The theory of diatomic molecules
 - 11 Scattering theory

- 12 The method of second quantization and radiation theory
- 13 Relativistic quantum mechanics
- 14 Some problems of quantum electrodynamics
- 15 Fundamentals of the theory of elementary particles

Subject index

Volume 4 (for details see p. xv)

Part VI Quantum Statistics and Physical Kinetics

- Chapter 1 Quantum statistics
- 2 Physical kinetics
- 3 The kinetic theory of gases and gas-like systems
- 4 The time correlation function method and Onsager's theory
- 5 Solid-state theory
- 6 The kinetic properties of solids
- 7 Interaction of radiation with a free-electron gas

Subject index



Contents of Volume 4

Part VI Quantum statistics and physical kinetics

Chapter 1 Quantum statistics

§ 1	The density matrix and the statistical operator	3
2	The statistical distribution in quantum statistics	7
3	Statistical distributions in an ideal gas	15
4	Degenerate ideal Bose gas	18
5	Imperfect Bose gas. Superfluidity	24

Chapter 2 Physical kinetics

6	Phenomenological transport equations and the general equations of kinetics	33
7	The mass conservation law and diffusion flow	36
8	The momentum conservation law and the equations of motion for a continuous medium	40
9	The energy conservation law and entropy transport in a moving continuous medium	44
10	The Fokker—Planck equation	50
11	The basic kinetic equation	58
12	Discussion of the basic kinetic equation and some simple examples	64
13	Non-equilibrium systems with a negative temperature. The amplification of electromagnetic waves by such systems	68

Chapter 3 The kinetic theory of gases and gas-like systems

14	Boltzmann's kinetic equation	72
15	The basic kinetic equation for correlation functions	79
16	The derivation of Boltzmann's equation from the basic kinetic equation	83

§17	The generalized transport equation and the properties of summational (additive) invariants	90
18	The equations of motion of a continuous medium	93
19	The laws of increase of entropy	99
20	Equilibrium and local-equilibrium distributions in an ideal gas	101
21	The general theory of the solution of Boltzmann's equation	106
22	The equations of hydrodynamics. The viscosity and thermal conductivity of gases	113
23	Relaxation time	123
24	The diffusion of an admixture of a gas of light particles into a gas of heavy particles	126
25	Thermal diffusion in gases	133
26	The dispersion of sound	136
27	The linearized Boltzmann's equation for quasi-gaseous systems	138
28	The solution of Boltzmann's equation for a quasi-gaseous system in an external force field	142
29	The kinetic equation for polyatomic gases	145
30	The moderation of fast neutrons	151
31	The spatial distribution of neutrons	161
32	Kinetic equation for a plasma disregarding collisions	166
33	The dispersion and damping of plasma waves	169
34	The kinetic equation for a plasma taking into account collisions	179
35	The establishment of equilibrium in an electron-ion plasma	185

Chapter 4 The time correlation function method and Onsager's theory

36	The response of a system to an external dynamic perturbation. Classical calculation	190
37	The response of a system to an external dynamic perturbation. Quantum calculation	197
38	The response of a system to a thermal perturbation	199
39	The calculation of kinetic coefficients. The connection with Boltzmann's equation	204
40	Onsager's theory	208
41	Discussion of Onsager's relations	213
42	Non-equilibrium processes in one-component systems	215

§43 Non-equilibrium processes in many-component systems (diffusion, thermodiffusion, thermoelectric effects)	220
44 Fluctuation-dissipation theorem	225

Chapter 5 Solid-state theory

45 A solid body as a quantum-mechanical system	229
46 The crystal lattice	232
47 Lattice vibrations	235
48 Wave function of an electron moving in a periodic field	241
49 The energy spectrum of an electron moving in a periodic field	244
50 A system of electrons in a solid	260
51 Models of a metal, a semiconductor and a dielectric	269
52 Magnetic properties of metals. The paramagnetism of an electron gas	271
53 The diamagnetism of an electron gas	275
54 Ferromagnetism	281
55 The interaction of electrons with lattice vibrations	287
56 The total Hamiltonian of a solid	294

Chapter 6 The kinetic properties of solids

57 The kinetic equation for electrons in metals	299
58 The electrical conductivity of metals	302
59 The Hall effect	305
60 The optical properties of a system of conduction electrons	310
61 The photoelectric effect	312
62 The mean free path length of electrons in metals	317
63 The collision integral for electrons in a metal	326
64 A solution of the kinetic equation	329
65 Superconductivity	334
66 Theory of the Fermi fluid	344
67 Electrons in dielectric crystals	351
68 The energy spectrum and the distribution function of electrons in semiconductors	355
69 The electrical conductivity and the Hall effect in semiconductors	362

*Chapter 7 Interaction of radiation with a free-electron gas**

§ 70	Low-density plasma in a low-frequency radiation field	365
71	Kinetic equations for electrons and photons	371
72	Kinetics of Bose condensation in a photon gas	376
73	Mobility of an electron in a radiation field	380
74	System of electrons in an arbitrary radiation field	386
75	General discussion of the results and range of the applicability of the theory	388
	Subject index	395

* Translated by Minerva Translations Limited, London.

PART VI

QUANTUM STATISTICS
AND PHYSICAL KINETICS

Quantum Statistics

§1. The density matrix and the statistical operator

In this chapter we shall consider some problems of statistical physics associated with the use of the concepts of quantum mechanics.

A system in quantum mechanics was assumed to consist of a relatively limited number of particles. However, as in classical physics, it is essential to carry out the transition to systems with very large numbers of particles. In other words, it is necessary to find a way of passing from quantum mechanics to quantum statistics.

In treating statistical physics in Part III, we took into account from the very beginning those profound changes which are brought about by quantum phenomena. That is, in calculating the state function the summation was carried out over the energy levels of the system, and the statistical weight of states was determined on the basis of the assumption that to each state there corresponds a phase-space cell of volume $(2\pi\hbar)^3$. In other words, we were, from the very beginning, constructing statistical physics in the quasi-classical approximation. Furthermore, we have already seen that the identity of quantum particles leads to a fundamental change in the properties of their statistical ensembles.

In quantum mechanics a substantiation of the quasi-classical approximation was given and, through this, the basic propositions on which our exposi-

tion of statistical physics was constructed were to a certain degree justified.

We are now going to pass on to a more logical construction of quantum statistical physics.

Let us consider a certain subsystem interacting with a medium. The medium and subsystem together form a closed system whose behaviour is described by a wave function Ψ . This wave function depends on the state of the particles of the subsystem as well as on the state of the particles of the medium. Since the subsystem interacts with the medium, it is impossible to find a wave function of the subsystem which depends only on the coordinates of its particles and not on those of the particles of the medium as well. Hence one cannot assign an independent wave function to an open subsystem. In other words, the states of an open subsystem are a set of mixed states. This is often formulated briefly as follows: 'an open subsystem is in a mixed state'.

As always in statistical physics, we shall be interested in the values of quantities L averaged over time (or over an ensemble). In order not to confuse this mean with the ordinary quantum-mechanical average, we shall denote the statistical mean by the symbol $\langle L \rangle$. We identify the quantity L with the operator \hat{L} . Its quantum-mechanical mean is

$$\overline{L} = \int \Psi^* \hat{L} \Psi d\tau. \quad (1.1)$$

Here \hat{L} and \overline{L} refer to the quasi-closed subsystem and depend only on quantities describing it, whereas the integration is carried out over the entire closed system.

Let us now consider our quasi-closed subsystem. We switch off its interaction with the external medium completely. Our subsystem then becomes closed. The Hamiltonian of the quasi-closed subsystem, $\hat{H} + \hat{H}_{\text{int}}$, where \hat{H}_{int} describes its interaction with the medium, goes over into the Hamiltonian \hat{H} . For clarity, we shall call such a small closed system a closed subsystem. It is obvious that a closed subsystem possesses properties completely different from those of a quasi-closed subsystem. Its states do not depend on the state of the external medium and, consequently, are pure states. A closed subsystem can be described by a particular wave function. If we denote the set of quantum numbers characterizing the state of the closed subsystem by n , then the set of wave functions ψ_n of this subsystem forms a complete set of orthonormalized functions. Hence we can write the expansion

$$\Psi = \sum c_n \psi_n. \quad (1.2)$$

The coefficients c_n depend on variables characterizing the state of the particles of the external medium and on the time. However, they can always be

considered to satisfy the normalization condition

$$\sum |c_n|^2 = 1. \quad (1.3)$$

Substituting the expansion (1.2) into the definition of the quantum-mechanical mean, we get

$$\bar{L} = \sum_{n,m} c_n^* c_m \int \psi_n^* \hat{L} \psi_m d\tau = \sum_{n,m} c_n^* c_m L_{nm}. \quad (1.4)$$

Formula (1.4) gives a complete quantum-mechanical description of the quantity \hat{L} . Its uncertainty, i.e. the fact that L can take on any value and that we can find only the mean value \bar{L} , is associated with the essence of the quantum-mechanical description. If the subsystem were closed, the coefficients c_n would not depend on time and $|c_n|^2$ would give the probability of finding the closed subsystem in the n th state. The coefficients c_n for an open subsystem do not possess such a property.

We now pass to the statistical mean, writing

$$\langle L \rangle = \sum \overline{c_n^* c_m}^t L_{nm}. \quad (1.5)$$

The average over time is carried out for a time large in comparison with microscopic times.

We now introduce a matrix determined by the set of elements $\overline{c_n^* c_m}^t$. The elements of this matrix are

$$\rho_{mn} = \overline{c_n^* c_m}^t. \quad (1.6)$$

Then (1.5) can be written in the form

$$\langle L \rangle = \sum_{n,m} \rho_{mn} L_{nm}. \quad (1.7)$$

Further, we introduce an operator $\hat{\rho}$ whose matrix elements are ρ_{nm} . The matrix ρ_{nm} is called the statistical matrix or density matrix, and the operator $\hat{\rho}$ the statistical operator. It is obvious that the last formula can then be written in the form

$$\langle L \rangle = \sum_n \left(\sum_m \rho_{mn} L_{nm} \right) = \sum_n (\hat{\rho} \hat{L})_{nn} = \text{Tr } \hat{\rho} \hat{L}. \quad (1.8)$$

We see that, in order to find the statistical mean of an operator \hat{L} , it is necessary to know the density matrix (or the statistical operator) replacing the distribution function of classical statistics.

In the statistical mean a further step is made in comparison with the quantum-mechanical mean: the interaction with the medium, which may have

a very complex character, is taken into account. Taking account of this interaction accurately is replaced by the averaging over time of the whole set of coefficients $c_n^* c_m \rightarrow \bar{c}_n^* \bar{c}_m$. It is clear that the statistical description (definition of values $\langle L \rangle$) is incomplete as compared with the quantum-mechanical description (i.e. definition of values \bar{L}).

In a description by means of a wave function it is possible, for example, to indicate possible accurate values of various quantities characterizing the system as a whole, even their unique possible values. In describing a system by means of the statistical matrix such predictions are impossible and we are forced to confine ourselves to the calculation of statistical means. In forming a statistical mean the particular features of the quantum-mechanical description are taken into account automatically. The situation here is exactly the same as in classical mechanics and statistics. In passing to a system with a very large number of particles, the dynamical description is replaced by a statistical one. It is just this latter which is adequate to describe the physical properties of the system.

We would like to warn the reader against the incorrect understanding of the statistical mean in quantum mechanics as a simple sequence of two averages; the quantum-mechanical one and the statistical one. To carry out quantum-mechanical averaging, according to the general formula, it is necessary to define the wave function. An open subsystem is in a mixed state and possesses no wave function. The averaging in formula (1.1) is carried out over the existing but unknown wave function of the closed system (subsystem + medium). It is clear that finding the wave function Ψ represents a problem of fantastic complexity.

The significance of formula (1.7) lies in the fact that to find statistical means it is not necessary to find either the wave function Ψ of the closed system or to describe in detail the behaviour of the quasi-closed subsystem.

To find the mean $\langle L \rangle$ it is necessary to know only the density matrix ρ_{mn} (or the statistical operator equivalent to it) and the matrix elements L_{nm} . The latter are calculated by means of the wave functions of the closed subsystem and are determined only by the properties of this subsystem.

The situation turns out to be very similar to that in classical statistics. In order to calculate means it is necessary to define the distribution functions and the multiplicity of degeneracy of the states of the subsystem without taking into account its interaction with the medium.

It should be stressed that so far we have by no means specified the properties of the quasi-closed subsystem. It can contain a large or a small number of particles. Only the fact that the subsystem is in a mixed state is essential. For any system in a mixed state the density matrix plays the role of a wave function for a system in a pure state.

§2. The statistical distribution in quantum statistics

Let us find the dependence of the statistical operator on time. For this we consider the mean value of the operator, $\langle \hat{L} \rangle$, assuming that the operator \hat{L} does not depend explicitly on time. It follows from (1.8) that

$$\langle \hat{L} \rangle = \text{Tr} \left(\frac{\partial \hat{\rho}}{\partial t} \hat{L} \right). \quad (2.1)$$

On the other hand, according to (31.2) of Part V*, we can write

$$\dot{\hat{L}} = \frac{i}{\hbar} [\hat{H}, \hat{L}], \quad (2.2)$$

where \hat{H} is the Hamiltonian of the closed subsystem to which the operator \hat{L} refers.

Substituting (2.2) into (1.8), we obtain

$$\begin{aligned} \langle \hat{L} \rangle &= \frac{i}{\hbar} \text{Tr} \hat{\rho} [\hat{H}, \hat{L}] = \frac{i}{\hbar} \{ \text{Tr} (\hat{\rho} \hat{H} \hat{L}) - \text{Tr} (\hat{\rho} \hat{L} \hat{H}) \} = \\ &= \frac{i}{\hbar} \{ \text{Tr} (\hat{\rho} \hat{H} \hat{L}) - \text{Tr} (\hat{H} \hat{\rho} \hat{L}) \} = \frac{i}{\hbar} \text{Tr} [\hat{\rho}, \hat{H}] \hat{L}. \end{aligned}$$

For the two expressions for $\langle \hat{L} \rangle$ to be the same, it is necessary to have the relation

$$\frac{\partial \hat{\rho}}{\partial t} = \frac{i}{\hbar} [\hat{\rho}, \hat{H}] = \frac{i}{\hbar} (\hat{\rho} \hat{H} - \hat{H} \hat{\rho}). \quad (2.3)$$

The derivative with respect to time of the statistical operator differs in sign from that of ordinary operators.

Formula (2.3) can be written in matrix form as follows:

$$\frac{\partial \rho_{mn}}{\partial t} = \frac{i}{\hbar} \sum_k (\rho_{mk} H_{kn} - H_{mk} \rho_{kn}). \quad (2.4)$$

Formula (2.3) (or the equivalent formula (2.4)) determines the development in time of the system described by the density matrix. It expresses one of the most general laws of nature. However, the explicit form of the operator $\hat{\rho}$ (or of the matrix ρ_{mn}) is unknown and cannot be determined only from eq. (2.3) without using additional information about the properties of the subsystem.

For the present, we shall confine ourselves to the case of stationary states,

* Volume 1 consists of Parts I and II, Volume 2 of Parts III and IV and Volume 3 of Part V.

where the density matrix does not change in time. Later on, in kinetics, we shall discuss the laws of changes of state in time.

Under stationary conditions, (2.4) gives

$$[\hat{\rho}, \hat{H}] = 0$$

or

$$\sum_k (\rho_{mk} H_{kn} - H_{mk} \rho_{kn}) = 0. \quad (2.5)$$

We see that the statistical operator commutes with the Hamiltonian and is a constant of the motion.

We now go over to the energy representation, where the basis functions ψ_n in expansion (1.2) are the eigenfunctions of the Hamiltonian of the isolated subsystem $\hat{H}\psi_n = E_n\psi_n$. In this case, only the diagonal matrix elements $H_{kn} = E_n\delta_{kn}$ are different from zero.

It follows from formula (2.5) that in this case

$$\sum_k (\rho_{mk} \delta_{kn} E_n - \rho_{kn} \delta_{km} E_m) = 0. \quad (2.6)$$

This relation shows that only the diagonal elements of the density matrix can differ from zero. If one returns to the definition (1.6), then this means that

$$\overline{c_n^* c_m} = \delta_{nm} \overline{|c_n|^2}. \quad (2.7)$$

This equality emphasizes once more the particular features of a description by means of the density matrix. The existing interaction always brings the subsystem into a mixed state and disturbs the interference between states which is characteristic of systems in a pure state.

Relation (2.7) is often called the random-phase condition. Indeed, if the coefficients c_n, c_m depend on time, then the phases of individual coefficients are completely uncorrelated. Their averaging gives (2.7).

It is convenient to denote the non-zero diagonal elements of the density matrix by

$$\rho_{nn} = \overline{c_n^* c_n} = w_n. \quad (2.8)$$

Then the formula for the statistical mean assumes the familiar form

$$\langle \hat{L} \rangle = \sum_n w_n L_{nn}. \quad (2.9)$$

Formula (2.9) shows that the quantities w_n represent the probabilities for the subsystem to be found in the n th energy state.

Our next problem is to find the probability distribution w_n .

Let us consider, first of all, the case of a closed system. In quantum mechanics there are no closed systems in the literal sense of the word. Every real system consisting of atoms undergoes an interaction with its environment. This can be, for example, an interaction with an electromagnetic field. Moreover, in §34 of Part V we have seen that the total energy of a closed system cannot have a definite value constant in time. Hence the states of a closed system containing a large number of particles can be considered to be mixed. The existing interactions, which do not change the energy of the system substantially, lead to violation of the interference between the states and to a random distribution of phases. The energy of a closed system can be considered to lie in an interval $\delta\epsilon_i \ll \epsilon_i$ but to constantly vary in this interval.

The hypothesis that a macroscopic system is in a mixed state is the basis of statistical physics. However, to find the explicit form of w_n it is necessary to put forward another hypothesis. Namely, it should be assumed that all the states of a closed system are equally probable. Hence it follows that w_n can be written in the form of the microcanonical distribution

$$w_n = \Omega(\epsilon_i). \quad (2.10)$$

We shall not reproduce here the reasoning of §15 and 16 of Part III, where the hypothesis of equal probabilities of states of a closed system was discussed in adequate detail. The derivation of the canonical distribution from the microcanonical one, given in §16 of Part III, is of general character.

In this derivation it was assumed only that the quasi-closed subsystem is a small part of the closed system described by the microcanonical distribution. The discrete character of the energy conditions has already been taken into account in the expression for the canonical distribution given in Part III.

In quantum statistics it is more convenient to write the normalized canonical distribution in the form

$$w_n = \frac{\exp(-\epsilon_n/kT)}{\sum \exp(-\epsilon_n/kT)} = \frac{\exp(-\epsilon_n/kT)}{Z}. \quad (2.11)$$

where the summation is carried out over all the states. The partition function (statistical sum) is of the form

$$Z = \sum w_n = \sum_n \rho_{nn} = \sum \exp(-\epsilon_n/kT). \quad (2.12)$$

Each degenerate state brings a number of terms equal to the multiplicity of the degeneracy into Z . For the statistical operator corresponding to the prob-

ability distribution (2.11) we obtain the expression

$$\hat{\rho} = Z^{-1} e^{-\hat{H}/kT}. \quad (2.13)$$

Indeed, if ψ_n is an eigenfunction of the operator \hat{H} , one can write the equality

$$\begin{aligned} \hat{\rho} \psi_n &= Z^{-1} e^{-\hat{H}/kT} \psi_n = Z^{-1} \left(1 - \frac{\hat{H}}{kT} + \frac{\hat{H}^2}{2(kT)^2} - \dots \right) \psi_n = \\ &= Z^{-1} \left(1 - \frac{\epsilon_n}{kT} + \frac{\epsilon_n^2}{2(kT)^2} - \dots \right) \psi_n = \\ &= Z^{-1} \exp(-\epsilon_n/kT) \psi_n = w_n \psi_n. \end{aligned} \quad (2.14)$$

The mean values and the partition function (statistical sum) can be written by means of the operator $\hat{\rho}$ as

$$\langle \hat{L} \rangle = \frac{\text{Tr } \hat{L} e^{-\hat{H}/kT}}{\text{Tr } e^{-\hat{H}/kT}}. \quad (2.15)$$

and

$$Z = \text{Tr } e^{-\hat{H}/kT}. \quad (2.16)$$

In calculating the trace, the summation is carried out over all the states of the subsystem.

We note that the operator form of notation of (2.13)–(2.15) does not depend on the choice of representation.

In the same way, without repeating the derivation of §59 of Part III, one can consider systems with a variable number of particles.

The statistical operator in the grand canonical distribution is of the form

$$\hat{\rho} = \frac{e^{(\mu \hat{n} - \hat{H})/kT}}{\text{Tr } e^{(\mu \hat{n} - \hat{H})/kT}}. \quad (2.17)$$

where \hat{n} is the operator of the number of particles, and μ is the chemical potential.

The grand partition function is

$$\tilde{Z} = \text{Tr } e^{(\mu \hat{n} - \hat{H})/kT} \quad (2.18)$$

or in matrix form

$$w_{n, \epsilon_i} = \tilde{Z}^{-1} \exp [(\mu n - \epsilon_i)/kT], \quad (2.19)$$

$$\tilde{Z}(V, T, \mu) = \sum_{n, i} \exp [(\mu n - \epsilon_i)/kT] = \sum_n z^n Z_n, \quad (2.20)$$

where n are the occupation numbers, z is the activity (see §59 of Part III), and Z_n is the partition function (2.12) for a subsystem made up of n particles, i.e.

$$z = e^{\mu/kT}, \quad Z_n = \sum_i \exp(-\epsilon_i/kT).$$

By means of the grand canonical distribution one can find mean values according to formula (1.18). In particular, the mean value of the number of particles in a subsystem

$$\langle N \rangle = \text{Tr } \hat{\rho} \hat{n} = \frac{\text{Tr} (e^{(\mu \hat{n} - \hat{H})/kT} \hat{n})}{\text{Tr} e^{(\mu \hat{n} - \hat{H})/kT}} = kT \frac{\partial}{\partial \mu} \ln \tilde{Z}. \quad (2.21)$$

According to §59 of Part III, the equation of state has the form

$$pV = kT \ln \tilde{Z}. \quad (2.22)$$

Hence we shall restrict ourselves to the case of an ideal gas confined in a container of volume V .

In the ideal gas the operator \hat{H} is just the kinetic energy operator

$$\hat{H} = \hat{H}_0 = \sum \frac{\hat{p}^2}{2m}. \quad (2.23)$$

Correspondingly, the normalized wave functions of free particles are of the form

$$\Psi_{\hat{p}}(\mathbf{r}) = V^{-1/2} e^{i(\mathbf{p} \cdot \mathbf{r})/\hbar}. \quad (2.24)$$

In a large volume the momenta of the particles can be considered to vary practically continuously.

The total symmetrized gas wave function (see §65 of Part V) is given by the general formula

$$\Psi_n = \left(\frac{n_1! n_2! \dots n_p! \dots}{N!} \right)^{1/2} \sum (-1)^p |\psi_1(\mathbf{r}_1) \dots \psi_{p_N}(\mathbf{r}_N)|. \quad (2.25)$$

It is obvious that Ψ is an eigenfunction of the operator \hat{H}_0 , i.e.

$$\hat{H}_0 \Psi = E \Psi, \quad (2.26)$$

where E is the energy of the entire gas

$$E = \sum \frac{p^2}{2m}. \quad (2.27)$$

The summation is carried out over all the permutations, the plus sign referring to bosons and the minus sign to fermions.

According to (2.15) and (2.23), the partition function assumes the form

$$Z = \text{Tr} e^{-\hat{H}_0/kT} = \sum \Psi^* e^{-\hat{H}_0/kT} \Psi. \quad (2.28)$$

Since Ψ is an eigenfunction of the operator \hat{H}_0 , one can write, analogously to (2.14),

$$e^{-\hat{H}_0/kT} \Psi = e^{-E/kT} \Psi.$$

so that

$$Z = \sum e^{-E/kT} |\Psi|^2. \quad (2.29)$$

The summation is carried out over all the states, i.e. over all possible values of the momenta of the gas particles. Since the momenta vary practically continuously, the summation can be replaced by integration, writing

$$\begin{aligned} Z &= \int e^{-E/kT} |\Psi|^2 \frac{d\Gamma}{(2\pi\hbar)^{3N}} = \\ &= \frac{V^N}{(2\pi\hbar)^{3N}} \int e^{-E/kT} |\Psi|^2 dp_1 \dots dp_{3N} = \\ &= \frac{V^N}{(2\pi\hbar)^{3N}} \int \exp\left(-\sum (p^2/2mkT)\right) |\Psi|^2 d\Gamma_p. \end{aligned} \quad (2.30)$$

To calculate the integral the simplest procedure is to write Ψ for a system of two particles. The generalization to N particles can then easily be made. Thus we calculate the auxiliary integral

$$\begin{aligned} I_2 &= \frac{1}{2} \int \exp\left(-\frac{p_1^2 + p_2^2}{2mkT}\right) |\psi_{\mathbf{p}_1}(\mathbf{r}_1) \psi_{\mathbf{p}_2}(\mathbf{r}_2) \pm \psi_{\mathbf{p}_1}(\mathbf{r}_2) \psi_{\mathbf{p}_2}(\mathbf{r}_1)|^2 d\mathbf{p}_1 d\mathbf{p}_2 = \\ &= \frac{1}{2} V^{-2} \int \exp\left(-\frac{p_1^2 + p_2^2}{2mkT}\right) \left| \exp\left(\frac{i(\mathbf{p}_1 \cdot \mathbf{r}_1 + \mathbf{p}_2 \cdot \mathbf{r}_2)}{\hbar}\right) \pm \exp\left(\frac{i(\mathbf{p}_2 \cdot \mathbf{r}_1 + \mathbf{p}_1 \cdot \mathbf{r}_2)}{\hbar}\right) \right|^2 d\mathbf{p}_1 d\mathbf{p}_2 = \\ &= \frac{1}{2} V^{-2} \int \exp\left(-\frac{p_1^2 + p_2^2}{2mkT}\right) \times \\ &\times \left[2 \pm \exp\left(\frac{i\mathbf{p}_1 \cdot (\mathbf{r}_2 - \mathbf{r}_1)}{\hbar}\right) \exp\left(\frac{-i\mathbf{p}_2 \cdot (\mathbf{r}_2 - \mathbf{r}_1)}{\hbar}\right) \mp \exp\left(\frac{i\mathbf{p}_1 \cdot (\mathbf{r}_2 - \mathbf{r}_1)}{\hbar}\right) \exp\left(\frac{i\mathbf{p}_2 \cdot (\mathbf{r}_2 - \mathbf{r}_1)}{\hbar}\right) \right] d\mathbf{p}_1 d\mathbf{p}_2 = \\ &= \frac{1}{2} V^{-2} \left[2(2\pi mkT)^3 \pm \int \exp\left(-\frac{p_1^2}{2mkT}\right) \exp\left(\frac{i\mathbf{p}_1 \cdot (\mathbf{r}_2 - \mathbf{r}_1)}{\hbar}\right) d\mathbf{p}_1 \int \exp\left(-\frac{p_2^2}{2mkT}\right) \exp\left(\frac{-i\mathbf{p}_2 \cdot (\mathbf{r}_2 - \mathbf{r}_1)}{\hbar}\right) d\mathbf{p}_2 \pm \right. \\ &\quad \left. \pm \int \exp\left(-\frac{p_1^2}{2mkT}\right) \exp\left(\frac{i\mathbf{p}_1 \cdot (\mathbf{r}_2 - \mathbf{r}_1)}{\hbar}\right) d\mathbf{p}_1 \int \exp\left(-\frac{p_2^2}{2mkT}\right) \exp\left(\frac{i\mathbf{p}_2 \cdot (\mathbf{r}_2 - \mathbf{r}_1)}{\hbar}\right) d\mathbf{p}_2 \right] = \\ &= V^{-2} (2\pi mkT)^3 \left[1 \pm \exp\left(-\frac{mkT|\mathbf{r}_2 - \mathbf{r}_1|^2}{2\hbar^2}\right) \right] = \\ &= V^{-2} (2\pi mkT)^3 \left[1 \pm \exp\left(-\frac{2\pi r_{12}^2}{\lambda_T^2}\right) \right]. \end{aligned} \quad (2.31)$$

with

$$\lambda_T = \left(\frac{2\pi\hbar^2}{mkT} \right)^{\frac{1}{2}}, \quad (2.32)$$

where λ_T is the wavelength of a particle of energy kT ('thermal wavelength' of a particle of mass m) and $r_{12} = |\mathbf{r}_2 - \mathbf{r}_1|$.

If we denote by f_{ik} the expression

$$f_{ik} = \exp \left(- \frac{\pi r_{ik}^2}{\lambda_T^2} \right) \quad (2.33)$$

then (2.31) can be written in the form

$$I_2 = V^{-2} (2\pi mkT)^3 [1 \pm f_{12} f_{21}]. \quad (2.34)$$

By analogy with (2.34), using (2.33) and substituting the value of Ψ from (2.25) into (2.30), we can easily write the general equation for the partition function

$$\begin{aligned} Z &= \frac{1}{N!} \left[\frac{2\pi mkT}{(2\pi\hbar)^2} \right]^{\frac{3}{2}N} \int \left[1 \pm \sum_{i < k} f_{ik}^2 + \sum_{i \neq k \neq l \neq m} f_{ik} f_{kl} f_{lm} \pm \dots \right] d\Gamma_V = \\ &= \frac{1}{N! (2\pi\hbar)^{3N}} \int e^{-E/kT} d\Gamma_p \int \left[1 + \sum_{i < k} f_{ik} + \dots \right] d\Gamma_V. \end{aligned} \quad (2.35)$$

The integration is carried out with respect to the coordinates of all the particles.

We see, first of all, that the expression for the partition function of an ideal gas is not the same as the quasi-classical expression

$$Z_{\text{qu-cl.}} = \left(\frac{2\pi mkT}{(2\pi\hbar)^2} \right)^{\frac{3}{2}N} \frac{V^N}{N!}$$

which we used in Part III, § 37.

If, however, the gas density is so small that the following inequality holds

$$\lambda_T \ll \bar{r}_{ik} \quad (2.36)$$

then it is seen from (2.33) that the quantities f_{ik} turn out to be very small in comparison with unity.

Neglecting the quantities f_{ik} in (2.35), we arrive at the equality

$$Z \approx Z_{\text{qu-cl.}} \quad (2.37)$$

It is easily seen that for atomic or molecular gases the condition (2.36) is much milder than that of neglecting the interaction between the particles.

Thus for not too dense atomic gases quantum corrections can always be disregarded.

In the case of dense gases these corrections are small in comparison with corrections for the fact that the gas is imperfect. Nevertheless, the character of quantum corrections to the quasi-classical expression for the partition function is of great theoretical interest.

Assuming the quantities f_{ik} to be small, we can, obviously, write

$$1 \pm \sum_{i < k} f_{ik}^2 \simeq \prod_{i < k} (1 \pm f_{ik}^2) = e^{-U(r_{ik}, T)/kT}. \quad (2.38)$$

$U(r_{ik}, T)$ is by definition equal to

$$U(r_{ik}, T) = -kT \ln (1 \pm f_{ik}^2) = -kT \ln [1 \pm \exp(-2\pi|\mathbf{r}_i - \mathbf{r}_k|^2/\lambda_T^2)]. \quad (2.39)$$

Substituting expression (2.38) into the partition function, we find

$$Z = (N!(2\pi\hbar)^{3N})^{-1} \int e^{-E/kT} d\Gamma_p \int \exp[-U(r_{ik}, T)/kT] d\Gamma_V. \quad (2.40)$$

We see that in our approximation, the quantum correction amounts to the appearance of an effective interaction between the particles. This interaction is characterized by the potential energy $U(r_{ik}, T)$ depending on the distance between the pairs of particles i and k , as well as on the temperature. This last fact shows at once that the function $U(r_{ik}, T)$ can only be considered as the effective but not the true potential energy.

Further, it should be stressed that the effective interaction is a purely quantum effect which disappears as $\hbar \rightarrow 0$.

We have discussed the existence of the effective interaction in §67 of Part V. This effective interaction is not associated with the appearance of new forces, but is due only to the symmetry of the wave function. In the case of bosons, $U(r_{ik}, T)$ is negative for all values of r_{ik} , i.e. the effective interaction has the character of an attraction. For fermions the effective interaction represents a repulsion, in accordance with what was said in §67 of Part V.

We now turn to the discussion of condition (2.36).

If we substitute into it the value of the thermal wavelength λ_T , then we can rewrite it in the form

$$\frac{2\pi\hbar}{(2\pi mkT)^{1/2}} \ll \left(\frac{V}{N}\right)^{1/3}. \quad (2.36')$$

Comparing this with formula (72.16) of Part III, we see that condition (2.36') represents none other than the condition for an ideal gas to be considered non-degenerate. Thus the criterion of degeneracy can be given an obvious meaning: an ideal gas can be considered non-degenerate if the thermal wavelength of the particles is small in comparison with the mean distance between them. Otherwise there is strong degeneracy. In this case the quantities f_{ik} turn out to be of the order of magnitude of unity, and the quantum effects no longer have the character of corrections but assume fundamental significance. In particular, the effective quantum interaction between particles determines to a considerable degree the properties of the system as a whole. We have seen this in Part III in the example of a degenerate Fermi gas. To avoid misunderstanding, we stress that in this case one can no longer make use of formula (2.39), since it was obtained on the basis of inequality (2.36).

It can be shown that an analogous situation holds in the case of systems of interaction particles*. However, for such systems the criterion of applicability of (2.36), i.e. the criterion of applicability of the quasi-classical approximation in statistical physics, has the form

$$\lambda_T \ll d, \quad (2.41)$$

where d is a certain effective distance of interaction between the particles. This criterion is fulfilled for gases and condensed systems of atomic and molecular particles (except for liquid helium II; see below).

§3. Statistical distributions in an ideal gas

In Part III we have already obtained statistical distributions in an ideal gas. For what follows, however, we shall need a somewhat different derivation of them, based on the grand canonical distribution. In the case of an ideal gas, one can write for the grand partition function

$$\begin{aligned} \tilde{Z}(V, T, z) &= \sum_{n=0}^{\infty} z^n Z_n = \sum_{n=0}^{\infty} z^n \sum_{n_k} \exp \left(- \sum \epsilon_k n_k / kT \right) \\ &= \sum_{n=0}^{\infty} z^n \sum_{n_k} \prod_k \exp(-\epsilon_k / kT) \end{aligned} \quad (3.1)$$

* K.Huang, *Statistical mechanics* (Wiley, New York, 1963).

since for an ideal gas the energy ϵ of the system is equal to

$$\epsilon = \sum \epsilon_k n_k,$$

where n_k is the number of particles in the k th state. It is obvious that the number of particles, n_k , obeys the condition

$$n = \sum n_k.$$

The double summation (over the number of particles in the system and over the number of particles in a given state, n_k) is equivalent to a single summation over all the independent values of n_k . This is most simply seen in the example of the particles of a Fermi gas, for which $n_k = 0, 1$. Obviously, in this case we have

$$\begin{aligned} Z(V, T, z) &= \sum_{n=0}^1 z^n \sum_{n_k} \prod_k (\exp(-\epsilon_k/kT))^{n_k} = \prod_k \sum_{n_k=0}^1 (z \exp(-\epsilon_k/kT))^{n_k} = \\ &= \prod_k (1 + z \exp(-\epsilon_k/kT)) = \prod_k (1 + \exp((\mu - \epsilon_k)/kT)). \end{aligned}$$

Analogously for bosons

$$\begin{aligned} Z(V, T, z) &= \sum_{n_1} \sum_{n_2} \dots \prod_k (ze^{-\epsilon_1/kT})^{n_1} (ze^{-\epsilon_2/kT})^{n_2} \dots = \\ &= \prod_k \sum_{n_1} (ze^{-\epsilon_1/kT})^{n_1} \sum_{n_2} (ze^{-\epsilon_2/kT})^{n_2} \dots = \\ &= \prod_k \frac{1}{1 - z \exp(-\epsilon_k/kT)} = \prod_k \frac{1}{1 - \exp((\mu - \epsilon_k)/kT)}. \end{aligned} \quad (3.2)$$

In the case of a Fermi gas

$$\sum_{n_k} (\exp((\mu - \epsilon_k)/kT))^{n_k} = 1 + \exp((\mu - \epsilon_k)/kT).$$

For the Bose gas $n_k = 0, 1, 2, \dots, N$, so that

$$\sum_{n_k} (\exp((\mu - \epsilon_k)/kT))^{n_k} = 1 - e^{(\mu - \epsilon_k)/kT}.$$

The grand partition function (or its logarithm, which is actually needed for the calculations) can be written in symmetric form

$$\ln \tilde{Z} = \sum_k \ln (1 \pm \exp((\mu - \epsilon_k)/kT)), \quad \text{where} \begin{cases} + & \text{for Fermi gas} \\ - & \text{for Bose gas} \end{cases} \quad (3.3)$$

All mean values can be expressed in terms of the grand partition function \tilde{Z} . The latter depends on the variables μ , V and T . Instead of the chemical potential it is always possible to introduce the total number of particles. Indeed, mean occupation numbers are given by formula (2.21)

$$\bar{n}_k = kT \frac{\partial \ln \tilde{Z}}{\partial \mu} = \frac{1}{\exp [(\epsilon_k - \mu)/kT] \pm 1}, \quad \text{where } \begin{cases} + \text{ for Fermi gas} \\ - \text{ for Bose gas} \end{cases} \quad (3.4)$$

The chemical potential μ is determined by the relation

$$N = \sum \bar{n}_k = \sum \frac{1}{\exp [(\epsilon_k - \mu)/kT] \pm 1}, \quad (3.5)$$

which allows one, in principle, to find the dependence of μ on T , V and N .

In the case of the Bose gas, condition (3.5) requires that the chemical potential μ be negative. For the Fermi gas no restrictions are imposed upon the values of μ .

The most obvious thermodynamic characteristic is the equation of state of the gas. According to § 59 of Part III, the equation of state of the gas can be expressed in terms of \tilde{Z} in the form

$$pV = \mp kT \sum_k \ln (1 \mp \exp [(\mu - \epsilon_k)/kT]), \quad \text{where } \begin{cases} + \text{ for Fermi gas} \\ - \text{ for Bose gas} \end{cases}. \quad (3.6)$$

We have written all the relations without discussion, since this was given in Part III. We recall only that in actual use of the above relations for a macroscopic gas one can pass from summation to integration, writing

$$\sum_k \rightarrow g \int \frac{d\Gamma}{(2\pi\hbar)^3}; \quad \epsilon_k = \frac{\hbar^2 k^2}{2m},$$

where $g = (2s+1)$, and s is the spin of the particle. The transition to Boltzmann's statistics corresponds to the fulfillment of the condition

$$e^{\mu/kT} \ll 1. \quad (3.7)$$

Substituting the chemical potential of an ideal gas, one can write (3.7) in the form of (2.36) or in the form of (72.16) of Part III.

In Part III we dwelt in detail on the behaviour of an ideal Fermi gas at low temperatures. To avoid repetition, we shall now consider only some general problems.

Writing (3.6) in the form

$$pV = \pm \frac{4\pi g V kT}{(2\pi\hbar)^3} \int p^2 dp \ln (1 \pm e^{(\mu - \epsilon)/kT}) =$$

$$\begin{aligned}
&= \pm (gkT) 2\pi \left(\frac{2m}{(2\pi\hbar)^2} \right)^{\frac{3}{2}} V \int_0^\infty \epsilon^{\frac{1}{2}} d\epsilon \ln(1 \pm e^{(\mu-\epsilon)/kT}) = \\
&= \pm gkTV \frac{4\pi(2m)^{\frac{3}{2}}}{3(2\pi\hbar)^3} \int_0^\infty \frac{\epsilon^{\frac{3}{2}} d\epsilon}{e^{(\epsilon-\mu)/kT} \pm 1} = \frac{2}{3} E, \quad (3.8)
\end{aligned}$$

we see that between the pressure and energy of Fermi and Bose gases there exists the same relation as for the classical ideal gas.

At high temperatures, when condition (3.7) is fulfilled, one can expand the integrand in a series and write

$$\begin{aligned}
pV &\simeq g \frac{(\pi m)^{\frac{3}{2}} V(kT)^{\frac{3}{2}}}{(2\pi\hbar)^3} e^{\mu/kT} \int_0^\infty (1 \mp e^{(\mu-\epsilon)/kT}) \times \\
&\quad \times e^{-\epsilon/kT} \epsilon^{\frac{3}{2}} d\epsilon = \frac{g(\pi m)^{\frac{3}{2}} V(kT)^{\frac{3}{2}}}{(2\pi\hbar)^3} e^{\mu/kT} \left(1 \pm \frac{1}{4\sqrt{2}} e^{\mu/kT} \right). \quad (3.9)
\end{aligned}$$

In this formula the chemical potential μ refers to a Boltzmann gas (see (60.9) of Part III). Hence (3.9) can be rewritten in the form

$$pV = NkT \left(1 \pm \frac{\pi^{\frac{3}{2}} N \hbar^3}{2gV(mkT)^{\frac{3}{2}}} \right), \quad (3.10)$$

where the minus sign refers to the Fermi gas, and the plus sign to the Bose gas.

The corrections found for the pressure are, as was stressed above, small compared with the corrections associated with the interaction between particles.

In the case of fermions, as we have seen in Part III, the effect of repulsion, associated with the Pauli exclusion principle, assumes a decisive importance for a degenerate gas.

In the next section we shall discuss the manifestation of the effect of attraction in a degenerate Bose gas.

§4. Degenerate ideal Bose gas*

The properties of a degenerate Bose gas differ radically from those of a Fermi gas. This difference is associated with the absence of any restrictions upon the accumulation of particles in the zero-energy state in the Bose gas.

* We follow the treatment in the book of K. Huang, *Statistical mechanics* (Wiley, New York, 1963) where the reader can find a number of additional data.

The mean number of particles in the zero-energy state can, according to (3.4), be written in the form

$$\overline{n(0)} = \frac{1}{e^{-\mu/kT} - 1}. \quad (4.1)$$

We see that the value of $\overline{n(0)}$ is determined by the behaviour of $\mu(T)$ as $T \rightarrow 0$. As we have seen in the preceding section, the chemical potential for the Bose gas is always less than zero. At high temperatures it decreases in absolute value with decreasing temperature.

The chemical potential is defined by the formula (3.5) for the total number of particles. This formula can be written in the form

$$\begin{aligned} N = \overline{n(0)} + N' &= \frac{\Omega(0)}{e^{-\mu/kT} - 1} + \sum' \frac{\Omega(\epsilon_i)}{\exp [(\epsilon_i - \mu)/kT] - 1} = \\ &= \frac{1}{e^{-\mu/kT} - 1} + \sum' \frac{\Omega(\epsilon_i)}{\exp [(\epsilon_i - \mu)/kT] - 1}. \end{aligned} \quad (4.2)$$

The first term represents the mean number of particles in the zero-energy state, and the second term the number of particles in all the other states. The statistical weight of the zero-energy state is equal to unity (see §35 of Part III).

The second term can be written in the form

$$N' = I(\mu, V, T) = \int_0^\infty \frac{d\Gamma}{(2\pi\hbar)^3} \frac{1}{e^{(\epsilon - \mu)/kT} - 1}.$$

Hence the number of particles in the zero-energy state is

$$\overline{n(0)} = N - I(\mu, V, T). \quad (4.3)$$

Formula (4.2) allows one to express the chemical potential in terms of temperature and density, while formula (4.3) relates the number of particles in the zero-energy state and in states with $\epsilon \neq 0$ to the total number of particles.

We see, first of all, that if the value of the integral I on the right-hand side is bounded by a certain value N_{\max} , then the number of particles in the zero-energy state

$$\overline{n(0)} = N - N_{\max} > 0$$

turns out to be finite and to represent a certain fraction of the total number of particles.

Let us consider the quantity $I(\mu, V, T)$. It is obvious that it can be written

in the form

$$\begin{aligned}
 I(\mu, V, T) &= V \frac{(2\pi m)^{\frac{3}{2}}}{(2\pi\hbar)^3} \frac{2}{\pi^{\frac{1}{2}}} \int_0^{\infty} \frac{\epsilon^{\frac{1}{2}} d\epsilon}{e^{(\epsilon-\mu)/kT}} = \\
 &= V \frac{(2\pi m)^{\frac{3}{2}}}{(2\pi\hbar)^3} \frac{2}{\pi^{\frac{1}{2}}} \int_0^{\infty} \epsilon^{\frac{1}{2}} d\epsilon \sum_{l=1}^{\infty} e^{l\mu/kT} e^{-l\epsilon/kT} = \\
 &= V \frac{(2\pi m)^{\frac{3}{2}}}{(2\pi\hbar)^3} \frac{2}{\pi^{\frac{1}{2}}} \sum_{l=1}^{\infty} e^{l\mu/kT} \int_0^{\infty} e^{-l\epsilon/kT} \epsilon^{\frac{1}{2}} d\epsilon = \\
 &= V \frac{(2\pi m)^{\frac{3}{2}}}{(2\pi\hbar)^3} (kT)^{\frac{3}{2}} \sum_{l=1}^{\infty} \frac{e^{l\mu/kT}}{l^{\frac{3}{2}}} = V \frac{(2\pi mkT)^{\frac{3}{2}}}{(2\pi\hbar)^3} F\left(-\frac{\mu}{kT}\right). \quad (4.4)
 \end{aligned}$$

Thus the value of l is mainly determined by the sum

$$F\left(-\frac{\mu}{kT}\right) = \sum_{l=1}^{\infty} \frac{e^{l\mu/kT}}{l^{\frac{3}{2}}}. \quad (4.5)$$

This sum converges for all values $\mu \leq 0$. The quantity $F(-\mu/kT)$ represents a monotonically decreasing function of its argument, and

$$F\left(-\frac{\mu}{kT}\right) \leq F(0) = \sum_{l=1}^{\infty} l^{-\frac{3}{2}} \cong 2.61.$$

Correspondingly,

$$I(\mu, V, T) \leq N'_{\max} = \frac{2.61 V(2\pi mkT)^{\frac{3}{2}}}{(2\pi\hbar)^3}. \quad (4.6)$$

Hence

$$\overline{n(0)} \geq N - \frac{2.61 V(2\pi mkT)^{\frac{3}{2}}}{(2\pi\hbar)^3}. \quad (4.7)$$

If the equality

$$N = N'_{\max} = 2.61 \frac{V(2\pi mkT)^{\frac{3}{2}}}{(2\pi\hbar)^3}, \quad (4.8)$$

or

$$kT_0 = \frac{1}{(2.61)^{\frac{2}{3}}} \frac{(2\pi\hbar)^2}{(2\pi m)} \left(\frac{N}{V}\right)^{\frac{2}{3}}, \quad (4.9)$$

holds, then $\overline{n(0)} = 0$.

At $T > T_0$ there are no particles in the zero-energy state. More precisely, the number of particles in this state is $\overline{n(0)} \sim 1$ and is negligible in comparison with the number of particles in excited states ($N' \sim N$). When a temperature lower than T_0 is reached, a finite number of particles $\overline{n(0)}$, representing a certain fraction of the total number of particles, appears in the zero-energy state. The distribution of particles assumes a qualitatively new form: a finite number of particles is in one (zero) energy state, while the remaining particles are practically continuously distributed over all excited states. This phenomenon is called the Bose–Einstein condensation, and the temperature T_0 is said to be the condensation temperature.

The explicit dependence of $\overline{n(0)}$ on T and N/V can be found if the dependence of the chemical potential on these quantities is established. As we have seen, the latter is defined by eq. (4.2). Its solution (for example, graphical) shows that for $T = T_0$ the chemical potential reduces to zero. The curve of the dependence $(-\mu/kT) = f(T/T_0)$ is given in fig. VI.1.

For temperatures lower than the condensation temperature T_0 , μ can be set equal to zero in eq. (4.2). We then find

$$\begin{aligned} n(0) &= N - I(\mu = 0, T, V) = N - N' = \\ &= N - \frac{2.61 V(2\pi mkT)^{\frac{3}{2}}}{(2\pi\hbar)^3} = N \left(1 - \left(\frac{T}{T_0} \right)^{\frac{3}{2}} \right) \end{aligned} \quad (4.10)$$

Formula (4.10) defines the state distribution of particles for $T < T_0$. We see that a fraction $N(T/T_0)^{\frac{3}{2}}$ of all the particles are in excited states,

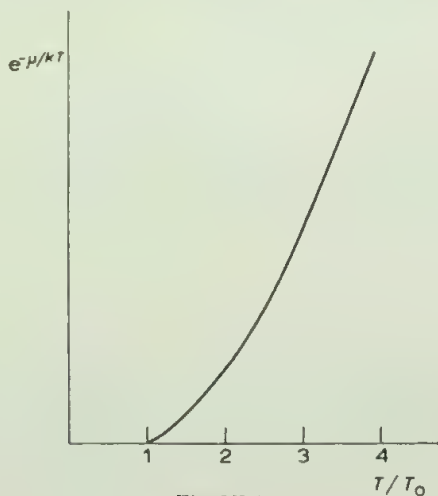


Fig. VI.1

while the remaining particles are condensed in the region of phase space corresponding to zero (or $\mathbf{p} = 0$).

The analogy with condensation becomes even more obvious if the expression for the pressure is determined.

According to §59 of Part III, we have

$$\begin{aligned}
 p &= \frac{kT}{V} \ln \tilde{Z} = -kTV^{-1} \sum \ln(1 - e^{-(\epsilon - \mu)/kT}) = \\
 &= -kTV^{-1} \left[\ln(1 - e^{\mu/kT}) + \sum' \ln(1 - e^{-(\epsilon - \mu)/kT}) \right] = \\
 &= -kTV^{-1} \ln(1 - e^{\mu/kT}) - kTV^{-1} \int' \frac{d\Gamma}{(2\pi\hbar)^3} \ln(1 - e^{-(\epsilon - \mu)/kT}).
 \end{aligned} \tag{4.11}$$

The prime means that the term with $\epsilon = 0$ is eliminated from the sum. In this formula we can set $V \rightarrow \infty$, since the volume of a macroscopic system is always large. Let us consider the regions $T \gg T_0$ and $T < T_0$. For $T \gg T_0$ and $V \rightarrow \infty$ the first term reduces to zero. In the second term one can write for μ the expression (60.9) of Part III for a Boltzmann gas, and we arrive at eq. (3.9). For $T < T_0$, as is shown by the analysis of eq. (4.2) for the chemical potential, the first term also reduces to zero as $V \rightarrow \infty$ and $\mu \rightarrow 0$.

Hence, integrating by parts, we find

$$p = \frac{4\pi(2m)^{\frac{3}{2}}}{3(2\pi\hbar)^3} \int_0^\infty \frac{\epsilon^{\frac{3}{2}} d\epsilon}{e^{\epsilon/kT} - 1} = \frac{0.08m^{\frac{3}{2}}(kT)^{\frac{5}{2}}}{(2\pi\hbar)^3}. \tag{4.12}$$

The pressure at $T < T_0$ turns out to be independent of density. Only particles in states with $\epsilon \neq 0$ give a contribution to the pressure. Since their number is $\sim T^{\frac{3}{2}}$, the total pressure turns out to be proportional to $T^{\frac{5}{2}}$.

The isotherms of the Bose gas have the form shown in fig. VI.2, where the dependence of the pressure on specific volume is plotted. At the point determined by condition (4.8), the specific volume is equal to

$$v_0 = \frac{V}{N} = \frac{(2\pi\hbar)^3}{2.61(2\pi mkT)^{\frac{3}{2}}}. \tag{4.13}$$

As can be seen from fig. VI.2, for $v < v_0$ the pressure is constant, whereas for $v > v_0$ it decreases with increasing specific volume.

The analogy between the curve of fig. VI.2 and the analogous isotherm for a liquid-vapour system is obvious. The region $v < v_0$ corresponds to the condensed phase, and the region $v > v_0$ to the vapour phase. The analogy with a phase transition is supplemented by the fact that at the condensation

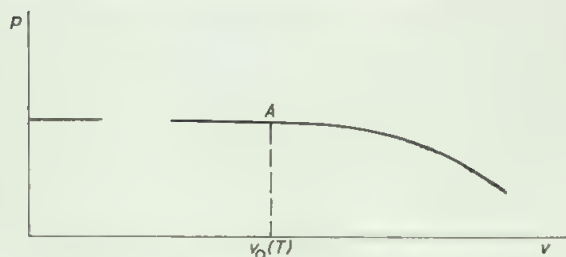


Fig. VI.2

point $T = T_0$ or $v = v_0$ latent heat of transition is released.

We shall not dwell on the calculations, but shall confine ourselves to two remarks. First, in addition to the similarity, it is necessary to stress the difference between the processes of condensation of ordinary vapour and the phenomenon of the Bose condensation. Ordinary condensation and transition into the liquid state are due to the interaction between molecules.

The Bose condensation is a process occurring in an ideal gas. Its nature is quite different and it is associated only with the quantum effect of the accumulation of bosons in the zero-energy state.

The second remark concerns the problem of the realization of the phenomenon of the Bose condensation. It is clear that since this phenomenon can occur only at low temperatures or large densities, the interaction between particles, which we have disregarded, must play an important role in its realization.

Liquid helium II, the only liquid which does not crystallize at low temperatures, displays a phase transition (see §77 of Part III). The ^4He particles have spin zero and obey Bose–Einstein statistics. However, it is not established whether this phase transition in the liquid is related to the phenomenon of the Bose–Einstein condensation in an ideal gas. As we shall see in the next section, the interaction between particles in an imperfect Bose gas is of basic importance. It turns out that the fundamental difference in the behaviour of ideal and non-ideal Bose gases is associated with this interaction. Hence, attempts at interpreting the phase transition in liquid helium as a manifestation of the Bose–Einstein condensation are somewhat arguable. Therefore, notwithstanding the theoretical importance of the phenomenon of the Bose–Einstein condensation, it is not certain that it occurs in nature.

In conclusion we stress that all the aforesaid referred only to Bose systems with a fixed number of particles. The phenomenon of the Bose–Einstein condensation cannot occur in Bose systems in which the number of particles is indefinite. Examples of such systems are photons or phonons. It is clear that

since the number of particles in such systems has no definite value, one cannot apply to them formula (4.2) which was assumed as the basis of the overall theory of the Bose-Einstein condensation.

In this chapter we shall confine ourselves to the consideration of bosons. In the chapter devoted to solid-state theory we shall consider the properties of Fermi systems in more detail.

§5. Imperfect Bose gas. Superfluidity

We now pass on to the consideration of an imperfect gas of bosons.

We shall assume that a certain sufficiently weak interaction exists between the particles of the Bose gas. It may have the character either of an attractive or a repulsive force.

The Hamiltonian of the system of particles in the x -representation is of the form

$$\hat{H} = - \sum_{\alpha=1}^N \frac{\hbar^2}{2m} \nabla_{\alpha}^2 + \frac{1}{2} \sum U(|\mathbf{r}_{\alpha} - \mathbf{r}_{\beta}|),$$

where N is the total number of particles in the system, and $U(|\mathbf{r}_{\alpha} - \mathbf{r}_{\beta}|)$ is the energy of interaction between the particles α and β . For a rarefied gas one need take into account only pair interactions.

For what follows it is convenient to consider the gas to be confined in a cube of edge L . Then the Hamiltonian of a free particle

$$\hat{H}_{\alpha} = - \frac{\hbar^2}{2m} \nabla_{\alpha}^2$$

will have a discrete spectrum. The momentum of a free particle will have components running over a discrete series of values

$$p_x = \frac{2\pi\hbar}{L} n_x, \quad p_y = \frac{2\pi\hbar}{L} n_y, \quad p_z = \frac{2\pi\hbar}{L} n_z,$$

where n_x, n_y, n_z are arbitrary integers, including zero. The eigenfunctions of the operator \hat{H}_{α} , normalized to a volume $V = L^3$, are of the form

$$\psi(\mathbf{r}) = V^{-1/2} e^{(i/\hbar)\mathbf{p} \cdot \mathbf{r}}.$$

We now apply the method of second quantization to the system of interacting particles. The Hamiltonian of a system of particles with pair interac-

tions is given by formula (99.25) of Part V

$$\hat{H} = \sum E(\mathbf{p}_k) \hat{\alpha}_{\mathbf{p}_k} \hat{\alpha}_{\mathbf{p}_k}^\dagger + \frac{1}{2} \sum_{i,k,l,m} (lm|U|ki) \hat{\alpha}_{\mathbf{p}_l}^\dagger \hat{\alpha}_{\mathbf{p}_m}^\dagger \hat{\alpha}_{\mathbf{p}_k} \hat{\alpha}_{\mathbf{p}_i} = \hat{H}_0 + \hat{H}'. \quad (5.1)$$

The summation is carried out over all discrete values of momenta (both positive and negative).

The energy $E(\mathbf{p}_k)$ represents the energy of a free particle

$$E(\mathbf{p}_k) = \mathbf{p}_k^2 / 2m.$$

By virtue of what has already been said the energy of interaction U involved in the matrix element of the Hamiltonian depends only on the coordinates of the two particles.

To calculate the matrix element use can be made of the wave functions of a free particle. This gives

$$\langle lm|U|ik \rangle = V^{-2} \int_V e^{-(i/\hbar)(\mathbf{p}_l \cdot \mathbf{r}_1 + \mathbf{p}_m \cdot \mathbf{r}_2)} U(|\mathbf{r}_1 - \mathbf{r}_2|) e^{(i/\hbar)(\mathbf{p}_i \cdot \mathbf{r}_1 + \mathbf{p}_k \cdot \mathbf{r}_2)} dV_1 dV_2.$$

Introducing new variables $\mathbf{q} = \mathbf{r}_1 - \mathbf{r}_2$ and $\mathbf{R} = \frac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)$ and integrating, taking into account that

$$\int_V e^{(i/\hbar)\mathbf{p} \cdot \mathbf{r}} dV = \begin{cases} V & \text{for } p = 0, \\ 0 & \text{for } p \neq 0, \end{cases}$$

we obtain

$$\langle lm|U|ik \rangle = \begin{cases} V^{-1} \nu(\mathbf{p}_l - \mathbf{p}_i) & \text{for } \mathbf{p}_l + \mathbf{p}_m = \mathbf{p}_i + \mathbf{p}_k, \\ 0 & \text{for } \mathbf{p}_l + \mathbf{p}_m \neq \mathbf{p}_i + \mathbf{p}_k, \end{cases} \quad (5.2)$$

where

$$\nu(\mathbf{p}) = \int U(|\mathbf{q}|) e^{-i\mathbf{p} \cdot \mathbf{q}} d\mathbf{q}.$$

Integration over all the angles can be carried out directly since U depends only on the absolute value of the vector \mathbf{q} . This gives

$$\begin{aligned} \nu(\mathbf{p}) &= \int U(|\mathbf{q}|) q^2 dq \int e^{-i\mathbf{p} \cdot \mathbf{q} \cos \theta} \sin \theta d\theta \int d\varphi = \\ &= 4\pi \int U(|\mathbf{q}|) \frac{\sin(pq)}{pq} q^2 dq. \end{aligned}$$

We see that the function $\nu(\mathbf{p})$ is real and that the equality $\nu(\mathbf{p}) = \nu(-\mathbf{p})$ is fulfilled. The operator of the total energy of the system of particles can, in

accordance with formulae (5.1) and (5.2), be expressed in the form

$$\hat{H} = \sum_{\mathbf{p}_k} \frac{\mathbf{p}_k^2}{2m} \hat{\alpha}_{\mathbf{p}_k}^\dagger \hat{\alpha}_{\mathbf{p}_k} + \frac{1}{2} \sum V^{-1} \nu(\mathbf{p}_l - \mathbf{p}_l) \hat{\alpha}_{\mathbf{p}_l}^\dagger \hat{\alpha}_{\mathbf{p}_m}^\dagger \hat{\alpha}_{\mathbf{p}_j} \hat{\alpha}_{\mathbf{p}_k}. \quad (5.3)$$

In the second term of formula (5.3) the sum is taken over only those values of the momenta $\mathbf{p}_i, \mathbf{p}_k, \mathbf{p}_l, \mathbf{p}_m$ for which the following relation holds

$$\mathbf{p}_l + \mathbf{p}_m = \mathbf{p}_i + \mathbf{p}_k.$$

The next problem consists in determining the eigenvalues of the Hamiltonian (5.3), i.e. in reducing the energy matrix to diagonal form.

Perturbation theory is inapplicable in studying the lowest energy levels. This is clearly seen from the fact that in the lower energy states, i.e. for small values of the momentum, the kinetic energy of the system tends to zero. On the other hand, the energy of interaction has a finite value and is therefore large in comparison with the kinetic energy. Therefore a special approximate method was developed by N.N. Bogoliubov for the investigation of weakly excited states. Let us first consider a system in which there is no interaction. In the energy ground state of such a system the momenta of all the particles are equal to zero, i.e.

$$\begin{aligned} n_{\mathbf{p}} &= 0 \quad \text{for } \mathbf{p} \neq 0, \\ n_{\mathbf{p}} &= N \quad \text{for } \mathbf{p} = 0, \end{aligned} \quad (5.4)$$

where $n_{\mathbf{p}}$ is the number of particles with momentum \mathbf{p} . It is natural to think that in the presence of a weak interaction in the lowest state of the system also, most particles will be found with momenta equal to zero.

In accordance with this we assume that

$$\begin{aligned} \hat{a}_0^\dagger \hat{a}_0 &= n_0 \approx N, \\ \hat{a}_{\mathbf{p}}^\dagger \hat{a}_{\mathbf{p}} &= n_{\mathbf{p}} \ll N. \end{aligned} \quad (5.5)$$

The operators \hat{a}_0^\dagger and \hat{a}_0 satisfy the commutation relation

$$\hat{a}_0 \hat{a}_0^\dagger - \hat{a}_0^\dagger \hat{a}_0 = 1.$$

Since $\hat{a}_0^\dagger \hat{a}_0 = n_0$ is large compared with one, we shall disregard the non-commutativity of these operators in what follows, i.e. we shall replace the operators \hat{a}_0 and \hat{a}_0^\dagger by ordinary numbers.

The assumptions (5.4) and (5.5) substantially simplify the expression for

the interaction Hamiltonian. Namely, in the sum over all momenta, only large terms, in which the factors \hat{a}_0^\dagger or \hat{a}_0 are involved in pairs or quadruply, need be retained. Such terms are

$$\hat{a}_0^\dagger \hat{a}_0^\dagger \hat{a}_0 \hat{a}_0, \quad \hat{a}_0^\dagger \hat{a}_0^\dagger \hat{a}_p \hat{a}_p, \quad \hat{a}_0^\dagger \hat{a}_p^\dagger \hat{a}_0 \hat{a}_p^\dagger$$

and so on. On the other hand, terms of the type $\hat{a}_0^\dagger \hat{a}_{p_m}^\dagger \hat{a}_{p_l} \hat{a}_{p_k}$, where $p_m \neq 0$, $p_l \neq 0$, $p_k \neq 0$, are small and can be dropped.

Thus for \hat{H}' we find

$$\begin{aligned} \hat{H}' &= \sum_{p_l + p_m = p_l + p_k} V^{-1} \nu(p_l - p_l) \hat{a}_{p_l}^\dagger \hat{a}_{p_m}^\dagger \hat{a}_{p_l} \hat{a}_{p_k} \simeq \\ &\simeq \sum_{p \neq 0} V^{-1} \nu(p) \{ \hat{a}_p^\dagger \hat{a}_0^\dagger \hat{a}_p \hat{a}_0 + \hat{a}_0^\dagger \hat{a}_p^\dagger \hat{a}_0 \hat{a}_p + \hat{a}_0^\dagger \hat{a}_0^\dagger \hat{a}_p \hat{a}_{-p} + \hat{a}_{-p}^\dagger \hat{a}_{-p}^\dagger \hat{a}_0 \hat{a}_0 \} + \\ &+ \nu(0) \hat{a}_0^\dagger \hat{a}_0^\dagger \hat{a}_0 \hat{a}_0 \simeq \\ &\simeq \nu(0) n^2 + \sum_{p \neq 0} V^{-1} \nu(p) [2 \hat{a}_p^\dagger \hat{a}_p n_0 + \hat{a}_p^\dagger \hat{a}_p^\dagger n_0 + \hat{a}_p \hat{a}_{-p} n_0] . \end{aligned}$$

Since the total number of particles in the system, N , is given by

$$N = n_0 + \sum_{p \neq 0} \hat{a}_p^\dagger \hat{a}_p ,$$

where the second term is small in comparison with the first, we have, in the same approximation

$$\begin{aligned} \hat{H}' &= \sum V^{-1} \nu(p_l - p_l) \hat{a}_{p_l}^\dagger \hat{a}_{p_m}^\dagger \hat{a}_{p_l} \hat{a}_{p_k} \simeq \\ &\simeq \nu(0) N^2 + 2N \sum_{p \neq 0} V^{-1} \nu(p) (\hat{a}_p^\dagger \hat{a}_p + \hat{a}_{-p}^\dagger \hat{a}_{-p}) + \\ &+ N \sum_{p \neq 0} V^{-1} \nu(p) (\hat{a}_p^\dagger \hat{a}_{-p} + \hat{a}_{-p}^\dagger \hat{a}_p) . \end{aligned} \quad (5.6)$$

The terms dropped are of the order of $N^{\frac{1}{2}}$.

To within a quantity of the order of N , the complete Hamiltonian \hat{H} assumes the form

$$\hat{H} = \sum_{p \neq 0} \left\{ \frac{p^2}{2m} + \frac{n_0}{V} \nu(p) \right\} \hat{a}_p^\dagger \hat{a}_p + \frac{1}{2} N^2 V^{-1} \nu(0) +$$

$$+ \frac{1}{2}NV^{-1} \sum_{\mathbf{p} \neq 0} \nu(\mathbf{p}) \hat{a}_{\mathbf{p}}^{\dagger} \hat{a}_{-\mathbf{p}}^{\dagger} + \frac{1}{2}NV^{-1} \sum_{\mathbf{p} \neq 0} \nu(\mathbf{p}) \hat{a}_{\mathbf{p}} \hat{a}_{-\mathbf{p}}. \quad (5.7)$$

We now introduce new operators $\hat{b}_{\mathbf{p}}^{\dagger}$ and $\hat{b}_{\mathbf{p}}$, which we define in the following way:

$$\hat{b}_{\mathbf{p}} = \frac{\hat{a}_0^{\dagger} \hat{a}_{\mathbf{p}}}{(n_0)^{\frac{1}{2}}}, \quad \hat{b}_{\mathbf{p}}^{\dagger} = \frac{\hat{a}_0 \hat{a}_{\mathbf{p}}^{\dagger}}{(n_0)^{\frac{1}{2}}}. \quad (5.8)$$

The operators $\hat{b}_{\mathbf{p}}$ and $\hat{b}_{\mathbf{p}}^{\dagger}$ defined in this way satisfy the same commutation relations as the operators $\hat{a}_{\mathbf{p}}$ and $\hat{a}_{\mathbf{p}}^{\dagger}$, since we consider \hat{a}_0 and \hat{a}_0^{\dagger} to be ordinary numbers. Furthermore, as is easily seen,

$$\hat{b}_{\mathbf{p}}^{\dagger} \hat{b}_{\mathbf{p}} = \hat{a}_{\mathbf{p}}^{\dagger} \hat{a}_{\mathbf{p}} = n_{\mathbf{p}}, \quad (5.9)$$

because $\hat{a}_0^{\dagger} \hat{a}_0 = n_0$. By means of operators $\hat{b}_{\mathbf{p}}$ and $\hat{b}_{\mathbf{p}}^{\dagger}$ the Hamiltonian (5.7) can be written in the form

$$\begin{aligned} \hat{H} = & \sum_{\mathbf{p} \neq 0} \frac{p^2}{2m} \hat{b}_{\mathbf{p}}^{\dagger} \hat{b}_{\mathbf{p}} + \frac{1}{2}N^2V^{-1} \nu(0) + \\ & + \frac{1}{2}n_0V^{-1} \sum_{\mathbf{p} \neq 0} \nu(\mathbf{p}) [\hat{b}_{\mathbf{p}}^{\dagger} \hat{b}_{-\mathbf{p}}^{\dagger} + \hat{b}_{\mathbf{p}} \hat{b}_{-\mathbf{p}} + 2\hat{b}_{\mathbf{p}}^{\dagger} \hat{b}_{\mathbf{p}}]. \end{aligned} \quad (5.10)$$

In order to bring the Hamiltonian (5.10) to diagonal form we carry out linear transformations to new operators $\hat{\xi}_{\mathbf{p}}$ and $\hat{\xi}_{\mathbf{p}}^{\dagger}$ by the substitutions

$$\hat{b}_{\mathbf{p}} = u_{\mathbf{p}} \hat{\xi}_{\mathbf{p}} + v_{\mathbf{p}} \hat{\xi}_{-\mathbf{p}}^{\dagger}, \quad \hat{b}_{\mathbf{p}}^{\dagger} = u_{\mathbf{p}} \hat{\xi}_{\mathbf{p}}^{\dagger} + v_{\mathbf{p}} \hat{\xi}_{-\mathbf{p}}, \quad (5.11)$$

where

$$u_{\mathbf{p}}^2 - v_{\mathbf{p}}^2 = 1; \quad u_{\mathbf{p}} = u_{-\mathbf{p}}; \quad v_{\mathbf{p}} = v_{-\mathbf{p}}.$$

If $u_{\mathbf{p}}$ and $v_{\mathbf{p}}$ are real functions of the momentum \mathbf{p} , then the new Bose operators $\hat{\xi}_{\mathbf{p}}$ and $\hat{\xi}_{\mathbf{p}}^{\dagger}$ satisfy the commutation relations (99.9) of Part V. Substituting (5.11) into (5.10) and requiring that the coefficients of operators of the type $\hat{\xi}_{\mathbf{p}}^{\dagger} \hat{\xi}_{\mathbf{p}}^{\dagger}$ and $\hat{\xi}_{\mathbf{p}} \hat{\xi}_{\mathbf{p}}$ reduce to zero, we find the functions $u_{\mathbf{p}}$ and $v_{\mathbf{p}}$. A simple but somewhat lengthy calculation gives

$$u_{\mathbf{p}} = \frac{1}{(1 - A_{\mathbf{p}}^2)^{\frac{1}{2}}}, \quad v_{\mathbf{p}} = \frac{A_{\mathbf{p}}}{(1 - A_{\mathbf{p}}^2)^{\frac{1}{2}}},$$

$$A_{\mathbf{p}} = \frac{V}{n_0 \nu(\mathbf{p})} \left\{ \epsilon(\mathbf{p}) - \frac{\mathbf{p}^2}{2m} - \frac{n_0}{V} \nu(\mathbf{p}) \right\},$$

$$\epsilon(\mathbf{p}) = \left(\frac{n_0 \mathbf{p}^2 \nu(\mathbf{p})}{Vm} + \frac{\mathbf{p}^4}{4m^2} \right)^{\frac{1}{2}}. \quad (5.12)$$

Then only diagonal terms are preserved in the operator \hat{H} , and it is brought to the form

$$\hat{H} = \hat{H}_0 + \sum_{\mathbf{p} \neq 0} \epsilon(\mathbf{p}) \hat{\xi}_{\mathbf{p}}^{\dagger} \hat{\xi}_{\mathbf{p}}. \quad (5.13)$$

The eigenvalues of the Hamiltonian (5.13) are

$$E = E_0 + \sum_{\mathbf{p} \neq 0} \epsilon(\mathbf{p}) n'(\mathbf{p}), \quad (5.14)$$

where

$$E_0 = \frac{1}{2} N^2 V^{-1} \nu(0) + \sum_{\mathbf{p} \neq 0} \frac{1}{2} \left[\epsilon(\mathbf{p}) - \frac{\mathbf{p}^2}{2m} - \frac{n_0}{V} \nu(\mathbf{p}) \right], \quad (5.15)$$

and the $n'(\mathbf{p})$ are integers.

One can also find the total momentum of a system of particles. It turns out to be equal to

$$\mathbf{P} = \sum_{\mathbf{p} \neq 0} \mathbf{p} \hat{a}_{\mathbf{p}}^{\dagger} \hat{a}_{\mathbf{p}} = \sum_{\mathbf{p}} \mathbf{p} \hat{b}_{\mathbf{p}}^{\dagger} \hat{b}_{\mathbf{p}}.$$

If in this equality we pass from the operators $\hat{b}_{\mathbf{p}}^{\dagger}$ and $\hat{b}_{\mathbf{p}}$ to the operators $\hat{\xi}_{\mathbf{p}}^{\dagger}$ and $\hat{\xi}_{\mathbf{p}}$, we obtain

$$\mathbf{P} = \sum_{\mathbf{p}} \mathbf{p} \hat{\xi}_{\mathbf{p}}^{\dagger} \hat{\xi}_{\mathbf{p}} = \sum_{\mathbf{p}} \mathbf{p} n'(\mathbf{p}). \quad (5.16)$$

Formulae (5.14) and (5.16) have a simple corpuscular interpretation. We see that the energy of the system is written in the form of the sum of two terms. The first term E_0 represents the energy of the ground (lowest) state. The second term can be interpreted as the energy of quasi-particles which represent the collective excitations of the system, $n'_{\mathbf{p}}$ is the number of elementary excitations in the state with momentum \mathbf{p} . The energy of each elementary excitation is equal to $\epsilon(\mathbf{p})$ (5.12).

For small momenta the energy of excitation can be written in the form

$$\epsilon(\mathbf{p}) = \left(\frac{\nu(0)}{m V_0} \right) |\mathbf{p}|. \quad (5.17)$$

Here $V_0 = V/n_0 \approx V/N$ represents the volume per particle. It follows from

formula (5.17) that the following inequality must be fulfilled

$$\nu(0) = \int U(q) dq > 0, \quad (5.18)$$

which means that repulsive forces are predominant. If, on the contrary, $\nu(0) < 0$, the energy turns out to be imaginary, which corresponds to an unstable state of the system.

For large momenta the energy is of the form

$$\epsilon(\mathbf{p}) = \frac{\mathbf{p}^2}{2m} + \frac{\nu(\mathbf{p})}{V_0}. \quad (5.19)$$

It should be noted that the existence of elementary excitations represents a collective effect of the entire system. Each excitation is associated with the state of the system as a whole, and not with the state of an individual particle. An ideal excitation gas, characterized by the Hamiltonian (5.13), obeys Bose statistics since the operators $\hat{\xi}_{\mathbf{p}}^\dagger$ and $\hat{\xi}_{\mathbf{p}}$ satisfy the commutation relations (99.9) of Part V.

It is easily shown that the system considered possesses the property of superfluidity. Suppose that the whole set of particles acquires an additional velocity v with respect to a certain system at rest, for example with respect to the walls of the tube or container containing the imperfect Bose gas. Then one can assume that all the results obtained will be valid for a system moving with velocity v with respect to a system at rest. If the energy in the moving system is equal to E^v , and that in the system at rest is equal to E , then the following relation exists between them

$$E^v = E + \frac{1}{2} N m v^2 + \mathbf{v} \cdot \mathbf{P}, \quad (5.20)$$

where \mathbf{P} is the total momentum of the Bose gas in the system at rest.

Using expressions (5.14) and (5.16), we obtain

$$E^v = E_0 + \frac{1}{2} N m v^2 + \sum n'(\mathbf{p}) \{ \epsilon(\mathbf{p}) + \mathbf{v} \cdot \mathbf{p} \}. \quad (5.21)$$

To decelerate the whole set of particles, there must arise excitations with momenta directed against the velocity v . The energy gain when one of these excitations appears is equal to

$$\Delta\epsilon = \epsilon(\mathbf{p}) - |\mathbf{v}| |\mathbf{p}|. \quad (5.22)$$

If the energy gain of the system is $\Delta\epsilon > 0$, the appearance of excitations is energetically unfavourable. This means that the system will move with the velocity v for an indefinitely long time without the appearance of such excitations. There is no deceleration in the system, and it possesses the property of superfluidity. Let us now formulate the condition for superfluidity. For the

quantity $\Delta\epsilon$ to be positive it is necessary that the following inequality be fulfilled

$$\frac{\epsilon(\mathbf{p})}{|\mathbf{p}|} > |\mathbf{v}|$$

for any \mathbf{p} . We denote the minimum value of the ratio $\epsilon(\mathbf{p})/|\mathbf{p}|$ by v^* . Then superfluidity occurs in motion with a velocity $v < v^*$. Consequently, superfluidity is possible in the system if the following inequality is fulfilled:

$$|\mathbf{v}^*| = \min \left(\frac{\epsilon(\mathbf{p})}{|\mathbf{p}|} \right) > 0. \quad (5.23)$$

It follows from the expression (5.12) for $\epsilon(\mathbf{p})$ that

$$v^* = \left(\frac{\epsilon(\mathbf{p})}{|\mathbf{p}|} \right)_{\mathbf{p} \rightarrow 0} = \left(\left[\frac{n_0 \nu(\mathbf{p})}{Vm} + \frac{\mathbf{p}^2}{4m^2} \right] \right)_{\mathbf{p} \rightarrow 0}^{\frac{1}{2}} = \left[\frac{n_0 \nu(0)}{mV} \right]^{\frac{1}{2}}.$$

If $\nu(0) > 0$, then v^* is real and positive. Thus in the case of the predominance of repulsive forces in an imperfect Bose gas there exists a real and positive value of v^* , and the Bose gas possesses the property of superfluidity.

In the case where there is no interaction between the particles (ideal Bose gas), $U(q) = 0$, $\nu(\mathbf{p}) = 0$ and the energy of the excitations is given by the formula

$$\epsilon(\mathbf{p}) = \mathbf{p}^2/2m. \quad (5.24)$$

Then $v^* = 0$, so that an ideal Bose gas possesses no superfluidity.

Thus we see that the property of superfluidity manifests itself in an imperfect Bose gas and is absent in an ideal Bose gas. Superfluidity is not associated with the specific properties of the system of bosons. For its existence a particular form of the energy spectrum of the collective excitations of the system is required. Namely, according to (5.23), it is necessary that the ratio of the minimum excitation energy of the system as a whole to the momentum of this excitation has a finite value.

The spectrum of excitations of a gas of non-interacting bosons has the form (5.24) for small excitations and does not satisfy (5.23).

The spectrum of excitations of a gas of interacting bosons satisfies condition (5.23) only in the presence of repulsive forces between the particles. Attractive forces do not lead to superfluidity for a system of interacting bosons. However, it should be stressed that in obtaining this result use was made of a property of bosons: we have assumed that the bulk of interacting particles is in the state with zero momentum, i.e. that it forms a condensate in momentum space. This is a necessary condition for the appearance of superfluidity in a Bose gas; however, this condition is not sufficient since the condensate also

forms in an ideal Bose gas but there is no superfluidity. The difference between the condensates of ideal and imperfect Bose gases is seen from the following reasoning. Let an ideal Bose gas move as a whole with a certain velocity v with respect to the walls of the container. If one of the particles is stopped as a result of an interaction with a wall of the container, the remaining gas particles will continue to move with a lower kinetic energy. Repetition of this process will in the end decelerate the gas.

The situation is different in the case of a gas whose particles undergo mutual repulsive forces. Stopping individual particles is in this case impossible. Interaction with the wall would excite the system as a whole. For a velocity of motion $v < v^*$ this turns out not to be possible.

It appears natural to apply this conclusion, obtained for an imperfect Bose gas, to a Bose fluid. Although the formal assumption of the pair character of the interactions is not fulfilled in a fluid, the qualitative discussion of the spectrum of small excitations of the collective motion is also applicable to a Bose fluid*.

* See K. Brueckner and K. Sawada, Phys. Rev. 106 (1957) 1117, 1128.

Physical Kinetics

§6. Phenomenological transport equations and the general equations of kinetics

The study of the behaviour of macroscopic systems which are not in a state of total thermodynamic equilibrium is as important in modern physics as consideration of the equilibrium states of macroscopic systems. Processes occurring in systems which are not in a state of total thermodynamic equilibrium are irreversible. The study of irreversible processes is the purpose and content of physical kinetics.

It is clear that the properties of non-equilibrium macroscopic systems, which are subjected to external actions and which develop in time, are immeasurably more complex than those of equilibrium systems. This can be seen from the infinite diversity of external actions which can violate the equilibrium in a system.

Hence it is not surprising that at present, physical kinetics has not reached in its development that degree of completeness and universality which is characteristic of statistical physics.

We need, first of all, to formulate certain general laws concerning changes of state of macroscopic systems which are not in an equilibrium state.

Two approaches to the solution of the problem are possible; a quasi-macroscopic one and a kinetic one.

In the quasi-macroscopic treatment the state of the system is defined by certain macroscopic parameters; temperature, concentration and so on. One seeks the laws of change of these parameters in space and time, when the system is not in an equilibrium state but is subjected to the action of external forces. In this case it is assumed that changes in the states of a macroscopic system can be characterized by the probability of transition from one macroscopic state into another. In other words, it is assumed that the concept of probable change of the macroscopic parameters characterizing a system can be introduced.

In the kinetic approach to the solution of the problems of physical kinetics it is assumed that the system can be described by a certain distribution function depending, in general, on the generalized coordinates and momenta of the particles. Assuming the probabilities of transition from one microstate to another (i.e. of the transition of a figurative point from one phase-space cell to another) to be defined, one works out the law of change of the distribution function, the so-called kinetic equation. The solution of this allows one to find the distribution function of the non-equilibrium system. It is clear that the kinetic approach is more detailed. In particular, it makes it possible to calculate quantities such as the kinetic coefficients; thermal conductivity, diffusion coefficient and so on.

Deferring the discussion of the kinetic approach to the study of non-equilibrium systems, we shall first discuss the quasi-macroscopic method. It is clear that it has a limited region of application: if a system is not in equilibrium, its state cannot, in general, be characterized by means of macroscopic parameters; for example, one cannot speak of the temperature or pressure of a body which is not in an equilibrium state. These concepts, as we have already seen, have a definite meaning only for a system placed in a reservoir. Nevertheless, in a number of cases an important property of macroscopic systems which we have already discussed, in §24 and §25 of Part III, allows us to make use of macroscopic quantities for describing the states of non-equilibrium systems.

We have seen that in a system which is not in an equilibrium state and is left to itself the phenomenon of relaxation occurs. After the lapse of the relaxation time the system goes over into an equilibrium state. In a real macroscopic system, the approach to an equilibrium state is very frequently accompanied by a number of processes. We shall illustrate this by a simple example.

Let two gases whose particles are of substantially different masses, for example gases of ions and electrons, exist in a container. The following sequence of processes occurs in the transition of the system into an equilibrium state:

(1) the gases in mixing fill the container uniformly; (2) statistical equilibrium is established in the gases, and then, because the mass difference hinders the transfer of momentum in collisions between the particles of the different gases, only collisions between identical particles are effective, as a result of which after the lapse of relaxation times τ_1 and τ_2 respectively, a Maxwell distribution is established in each of the gases separately; (3) as a result of momentum transfer in collisions between particles of different mass, after the lapse of a considerably longer relaxation time τ_3 , a common Maxwell distribution will be established in the mixture of gases; (4) capture of electrons by ions, accompanied by the emission of the corresponding energy, will, under certain conditions, lead to the transition of the gas into a neutral equilibrium state (time τ_4). Thus the approach of the gas to total equilibrium proceeds in several stages having relaxation times which differ substantially from each other.

The properties mentioned are possessed by a very large class of physical systems, in particular by any systems which consist of small parts. Since the relaxation time increases rapidly with increasing size of the system, equilibrium is established more rapidly inside the parts than between the parts (see §25 of Part III).

In practice, the study of systems in which an equilibrium with respect to fast processes is established during the relaxation time τ_{fast} , but in which an equilibrium for slowly varying parameters does not have time to be established, is often of great interest. Such systems are said to be in a state of incomplete equilibrium, or are quasi-equilibrium systems.

The mixture considered above is in incomplete equilibrium. If the time t which has elapsed since the moment of mixing satisfies the inequality

$$\tau_1, \tau_2 \ll t \ll \tau_3,$$

each of the parts in equilibrium, the electron gas and the ion gas, can be characterized by ordinary macroscopic parameters. However, as distinct from equilibrium systems, these parameters will vary slowly in time.

If the same gas mixture is not isolated, but is in an external force field, for example in an electric field, then its state will change under the action of the field. For a field constant in time, and for $t \gg \tau_1, \tau_2$ the two systems can also be assumed to be independent of time, but to be non-equilibrium systems. Indeed, a systematic motion (flux) of charges in the direction of the field arises in this system.

A time-independent non-equilibrium state of a system is called a stationary state. In such a state, as well as in a state of incomplete equilibrium, the system can be characterized by the values of macroscopic parameters.

It follows from what has been said that a quasi-macroscopic consideration of non-equilibrium systems is possible in the case where the processes occurring in them are slow.

A disadvantage of the reasoning given is the relative character of the notions of fast and slow processes. For actual systems, one is often able to carry out this division sufficiently clearly. It is obvious that a necessary condition for such a division is the requirement that $\tau_{\text{slow}} \gg \tau_{\text{fast}}$.

If the system is in a state of incomplete equilibrium, the thermodynamic concepts such as temperature, thermodynamic potentials and so on are applicable to it or to its macroscopic parts.

For example, one can speak of a system with pressure and temperature varying from point to point or varying in time.

In a macroscopic system which is not in equilibrium, displacement of its parts and mass transport often occur. However, in this case these parts can be characterized by thermodynamic parameters; density, pressure and temperature.

In what follows we shall formulate empirical laws of transport in such non-equilibrium systems, and then we shall go on to the derivation of theoretical relations characterizing the behaviour of non-equilibrium systems.

§7. The mass conservation law and diffusion flow

As we have already pointed out in §6, the violation of equilibrium in a system is often associated with a macroscopic motion of its parts. Let us divide our macroscopic system into small but still macroscopic elements, and let us assume that these elements are in a state of local equilibrium. This means that each element can be assigned ordinary thermodynamic characteristics; definite temperature, mean density and thermodynamic potentials.

We shall see below that such local equilibrium for small elements is established extremely rapidly in the simplest system (the ideal gas). A departure from the equilibrium state of the system as a whole, in particular mechanical motion of its parts, does not violate the local equilibrium in small elements. By making the assumption of local equilibrium in small elements and of non-equilibrium of the system as a whole, i.e. considering the system to be in incomplete equilibrium, one can formulate general laws of change of state of such a system. It is then necessary to take into account its internal motion.

We shall abstract the molecular structure of the system and assume it to be a continuous medium. This means that the velocity of displacement $\mathbf{V}(\mathbf{r}, t)$, which is a continuous function of coordinates \mathbf{r} and time t , is considered to be defined at each point of the system.

A continuous medium can be a liquid, a gas or a solid body, flowing under the action of applied forces. However, one deals most often with incompressible fluids and gases. Accordingly, the density, pressure, temperature and other thermodynamic characteristics of the medium will be assumed to be continuous functions of coordinates and time. In order to avoid misunderstanding, we stress that the dependence of thermodynamic quantities on coordinates and time is to be understood as a variation of local equilibrium characteristics. For example, the energy E of a certain small (but macroscopic) element of a system changes when it is displaced from a position with pressure p_1 and temperature T_1 to another position with pressure p_2 and temperature T_2 . However, at each position the relation between E , p and T has the equilibrium thermodynamic character.

The continuous medium approximation corresponds to a thermodynamic description of equilibrium systems. In this approximation we shall formulate general laws determining the motion of the medium, and mass transport and energy transfer in it. In this case, naturally, we shall have to make use of certain empirical relations. The latter will not have the general character which is possessed by the empirical laws (the first and second) of thermodynamics, because, as we have already stressed, non-equilibrium systems display a great diversity of properties depending on their actual structure and on the character of the processes.

One of the most important problems of macroscopic kinetics is the derivation of the transport equations for a continuous medium and finding the constants, called the kinetic coefficients, involved in these equations.

Such general transport equations are those of mass, momentum, energy and entropy transfer. Let us first formulate the mass transport law. In a system of n components, the law of conservation of mass of each of the components holds. If the mean macroscopic velocity of the α th component is denoted by \mathbf{v}_α , and its mass per unit volume by ρ_α , then the mass conservation law can be written in the form

$$\frac{\partial \rho_\alpha}{\partial t} + \nabla \cdot (\rho_\alpha \mathbf{v}_\alpha) = 0. \quad (7.1)$$

Here it is assumed that no chemical reaction takes place between the components. In the more general case, one should write on the right-hand side of (7.1) the rate of appearance or disappearance of the particles of the α th component per unit volume. Eq. (7.1) is just the mass transport equation.

The motion of a system is more conveniently characterized by the velocity of the centre of mass rather than by the velocities of the components. By definition, the velocity of the centre of mass \mathbf{v} is equal to

$$\mathbf{v} = \frac{\sum \rho_{\alpha} \mathbf{v}_{\alpha}}{\sum \rho_{\alpha}} = \frac{\sum \rho_{\alpha} \mathbf{v}_{\alpha}}{\rho}, \quad (7.2)$$

where ρ is the total density in the system.

Summing (7.1) over all the components, we easily find the relation between ρ and \mathbf{v} :

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \quad (7.3)$$

This is the so-called hydrodynamic equation of continuity, expressing the law of conservation of mass of the entire system. If the total density of the medium is constant, then $\partial \rho / \partial t = 0$ and instead of (7.3) one can write

$$\nabla \cdot \mathbf{v} = 0. \quad (7.4)$$

Media with constant density are said to be incompressible. In hydrodynamics it turns out that if the velocity of a continuous medium (a liquid or a gas) is small in comparison with the velocity of sound, such a medium can be considered incompressible.

We rewrite (7.1) in the form

$$\frac{\partial \rho_{\alpha}}{\partial t} + (\mathbf{v}_{\alpha} \cdot \nabla) \rho_{\alpha} = -\rho_{\alpha} (\nabla \cdot \mathbf{v}) - \nabla \cdot (\rho_{\alpha} (\mathbf{v}_{\alpha} - \mathbf{v}))$$

or

$$\frac{d\rho_{\alpha}}{dt} = -\rho_{\alpha} (\nabla \cdot \mathbf{v}) - \nabla \cdot \mathbf{j}_{\alpha}, \quad (7.5)$$

where \mathbf{j}_{α} denotes the so-called diffusion flux of the α th component,

$$\mathbf{j}_{\alpha} = \rho_{\alpha} (\mathbf{v}_{\alpha} - \mathbf{v}). \quad (7.6)$$

Formula (7.5) expresses the law of conservation of mass of the α th component. The vector \mathbf{j}_{α} shows to what degree the motion of particles of the α th component differs from the mean velocity of motion of the system as a whole. The law of conservation of mass of the α th component is usually rewritten by introducing the mass concentration

$$c_{\alpha} = \rho_{\alpha} / \rho. \quad (7.7)$$

Then instead of (7.5), taking into account (7.6), we obtain

$$\frac{dc_{\alpha}}{dt} = -\frac{1}{\rho} (\nabla \cdot \mathbf{j}_{\alpha}). \quad (7.8)$$

It is obvious that in a system of n components, $n - 1$ concentrations and $n - 1$ diffusion fluxes are independent, since

$$\sum_{\alpha} c_{\alpha} = 1 ; \quad \sum_{\alpha} j_{\alpha} = 0 . \quad (7.9)$$

Of course, in a one-component liquid there is no diffusion flow. In this case, the law of conservation of mass is expressed by formula (7.3).

For simplicity of notation, we shall consider a two-component system, in which one concentration and one diffusion flux are independent.

In the absence of external fields, equilibrium conditions are the constancy of the chemical potential and of the temperature in the system (see Part III). Diffusion flow in equilibrium is equal to zero.

We now consider a system in a non-equilibrium state. For sufficiently small departures of the system from the equilibrium state, it is natural to assume that the diffusion flux which arises is proportional to the gradient of the chemical potential μ :

$$\mathbf{j} = -\gamma \nabla \mu . \quad (7.10)$$

We assume that $\gamma > 0$. Here the minus sign shows that the flow which arises is directed from a higher towards a lower chemical potential. This corresponds to the requirement of a minimum chemical potential in the equilibrium state. Formula (7.10) represents, in essence, the first term of the expansion in powers of the quantity $\nabla \mu$, and is meaningless for large departures from the equilibrium state.

In this section we shall confine ourselves to the case of an isothermal system.

For an isothermal system one can write

$$j_{\alpha}^{(M)} = -\gamma \left(\frac{\partial \mu}{\partial c_{\alpha}} \right)_{p,T} \nabla c_{\alpha} - \gamma \left(\frac{\partial \mu}{\partial p} \right)_{T,c_{\alpha}} \nabla p = -\rho D_{\alpha} \nabla c_{\alpha} - \frac{k_p^{(\alpha)}}{p} \nabla p , \quad (7.11)$$

where two new coefficients are introduced: the molecular diffusion coefficient

$$D_{\alpha} = \frac{\gamma}{\rho} \left(\frac{\partial \mu}{\partial c_{\alpha}} \right)_{p,T} \quad (7.12)$$

and the barodiffusion coefficient

$$\frac{k_p^{(\alpha)}}{p} = \frac{\gamma (\partial \mu / \partial p)_{c_{\alpha}, T}}{\rho D_{\alpha}} . \quad (7.13)$$

As a rule, the barodiffusion coefficient is very small, and the second term of (7.11) can be neglected. Then the diffusion flux assumes the form

$$\mathbf{j}_\alpha^{(M)} = -\rho D_\alpha \nabla c_\alpha. \quad (7.14)$$

Instead of the mass concentration c_α , use is often made of the number of particles per unit volume, and the diffusion flux is related to the number of particles rather than to the mass. In this case

$$\mathbf{j}_\alpha = -D_\alpha \nabla c_\alpha. \quad (7.15)$$

Formulae (7.11), (7.14) and (7.15) represent the well-known empirical law of diffusion (Fick's law).

We shall see below that for non-isothermal systems the law of diffusion must be generalized somewhat. In what follows we shall see also that for the case of an ideal gas the diffusion law can be derived theoretically from the general laws of physical kinetics.

Taking into account (7.11), the mass conservation law (7.8) is written in the form

$$\frac{dc_\alpha}{dt} = D_\alpha \nabla^2 c_\alpha + k_p^\alpha \nabla \cdot \left(\frac{\nabla p}{\rho} \right) \quad (7.16)$$

or, disregarding barodiffusion,

$$\frac{\partial c_\alpha}{\partial t} + (\mathbf{v} \cdot \nabla c_\alpha) = D_\alpha \nabla^2 c_\alpha. \quad (7.17)$$

The last expression is called the equation of convective diffusion. It describes convective mass transport in a moving medium, as well as molecular diffusion.

§8. The momentum conservation law and the equations of motion for a continuous medium

In the preceding section we have considered the motion of a continuous medium to be defined. We now consider how to find the equations of motion of such a medium under the action of external forces applied to the system.

For this we write the equations of motion of a small (but finite) volume element dV of the medium in the form

$$\frac{d}{dt} \int \rho \mathbf{v} dV = \int \mathbf{F} dV + \oint \mathbf{F}^{(n)} dS, \quad (8.1)$$

where \mathbf{F} is the force per unit volume, and $\mathbf{F}^{(n)}$ is the surface force per unit area with which the surrounding medium acts upon a given volume. The vector $\mathbf{F}^{(n)}$ is called the stress. To define the stress, it is necessary to indicate the direction of the vector, \mathbf{n} , normal to the surface.

We introduce an elementary parallelepiped bounded by the planes $\mathbf{n}_1 = \mathbf{i}$, $\mathbf{n}_2 = \mathbf{j}$, $\mathbf{n}_3 = \mathbf{k}$. The vector $\mathbf{F}^{(i)}$ has, obviously, the components F_{xx} , F_{yx} , F_{zx} each of which represents a component of the force acting upon an area of 1 cm^2 with normal \mathbf{i} . The vector components $\mathbf{F}^{(j)}$ and $\mathbf{F}^{(k)}$ have analogous meanings.

We can introduce the stress tensor σ_{ik} , defining it by

$$\sigma_{ik} n_k = F_i^{(n)} \quad (8.2)$$

It is obvious that the components of the tensor σ_{ik} are the same as those of $\mathbf{F}^{(n)}$. We transform the surface integral into a volume integral by the Gauss-Ostrogradsky formula

$$\oint \mathbf{F}^{(n)} dS = \int \sigma_{ik} n_i dS = \int \frac{\partial \sigma_{ik}}{\partial x_k} dV.$$

We then have finally*

$$\frac{d}{dt} \int \rho v_i dV = \int \rho \frac{dv_i}{dt} dV = \int F_i dV + \int \frac{\partial \sigma_{ik}}{\partial x_k} dV. \quad (8.3)$$

Since the volume V is arbitrary, it follows from (8.3) that

$$\rho \frac{dv_i}{dt} = F_i + \frac{\partial \sigma_{ik}}{\partial x_k} \quad (8.4)$$

or, if the external volume forces are conservative, then

$$\rho \frac{dv_i}{dt} \equiv \rho \left(\frac{\partial v_i}{\partial t} + v_k \frac{\partial v_i}{\partial x_k} \right) = - \frac{\partial U}{\partial x_i} + \frac{\partial \sigma_{ik}}{\partial x_k}, \quad (8.5)$$

where U is the potential energy per unit volume. Eqs. (8.4) and (8.5) express the law of motion of a continuous medium.

Eq. (8.5) can be rewritten in another form by slightly transforming its left-hand side.

Namely, making use of the equation of continuity, we can write

$$\rho \frac{\partial v_i}{\partial t} = \frac{\partial(\rho v_i)}{\partial t} - v_i \frac{\partial \rho}{\partial t} = \frac{\partial}{\partial t} (\rho v_i) + v_i \frac{\partial}{\partial x_k} (\rho v_k).$$

* It is evident that $\int v_i (d\rho/dt) dV = 0$.

Hence, substituting for $\rho \partial v_i / \partial t$ from (8.5), we have

$$\frac{\partial}{\partial t} \rho v_i = - \frac{\partial}{\partial x_k} (\rho v_i v_k - \sigma_{ik}) - \frac{\partial U}{\partial x_i} \quad (8.6)$$

The left-hand side of (8.6) represents the rate of change of momentum per unit volume. The first term on the right-hand side represents the momentum flow density $j_{ik} = \rho v_i v_k - \sigma_{ik}$. Thus (8.6) expresses the momentum conservation law. The stress tensor σ_{ik} can be related to the velocities on the basis of experimental data.

For the equilibrium state, when the velocity of motion of the medium reduces to zero, one can write the condition of thermodynamic mechanical equilibrium:

$$F_i = \frac{\partial p}{\partial x_i} \quad (8.7)$$

Comparing (8.2) and (8.7), we arrive at the conclusion that for a liquid at rest

$$\sigma_{ik} = -p \delta_{ik} \quad (8.8)$$

As shown by experiment, irreversible processes arise in moving continuous media. It is impossible to write an expression for σ_{ik} valid for all moving media, liquids, gases and solid bodies, for any mode of flow.

We shall confine ourselves to the most fully investigated (although not in practice the most frequent) case of the so-called Newtonian fluids. For Newtonian fluids the stress tensor is a linear function of the velocity gradient.

When there is lamellar flow between two rigid walls, one of which is at rest and the other moving with velocity v , the surface of the solid body is acted upon by the stress $\sigma_{xy} = \eta v/L$, where L is the distance between the walls. It is easily seen that under these conditions $\sigma_{xy} = \text{const}$. Consequently, the same stress acts upon each cm^2 of an imaginary plane in the fluid. For an arbitrary law of motion, one can write a general expression for σ_{ik} , proceeding from the following two requirements:

(1) σ_{ik} is a linear function of the derivatives $\partial v_i / \partial x_k$ or $\partial v_k / \partial x_i$ ($i, k = 1, 2, 3$).

(2) For uniform rotation of the fluid as a whole, σ_{ik} reduces to its statistical expression, since no relative displacement of fluid layers occurs.

The combinations of derivatives which satisfy this condition are

$$\sigma_{ik} = \alpha \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right); \quad \beta \frac{\partial v_i}{\partial x_i} \quad (8.9)$$

It can be seen by a direct check that for $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$, where $\boldsymbol{\omega}$ is the angular velocity, these combinations of derivatives reduce to zero. Hence in the most general case one can write

$$\sigma_{ik} = -p\delta_{ik} + \alpha \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right) + \beta \frac{\partial v_j}{\partial x_j} \delta_{ik}.$$

This expression is usually written in the identical form

$$\sigma_{ik} = -p\delta_{ik} + \eta \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} - \frac{2}{3} \delta_{ik} \frac{\partial v_j}{\partial x_j} \right) + \zeta \delta_{ik} \frac{\partial v_j}{\partial x_j} = -p\delta_{ik} + \sigma'_{ik},$$

where

$$\eta = \alpha, \quad (8.10)$$

and

$$\zeta - \frac{2}{3}\eta = \beta. \quad (8.11)$$

The kinetic coefficients η and ζ are called the first (or shear) and second (or bulk) coefficients of viscosity.

For an incompressible fluid $\partial v_j / \partial x_j = 0$, and the stress tensor assumes the form

$$\sigma_{ik} = -p\delta_{ik} + \eta \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right). \quad (8.12)$$

Newton's law of viscosity holds for gases and certain liquids (above all for water).

The viscosity is a function of temperature

$$\eta \sim \begin{cases} T^{\frac{1}{2}} & \text{for gases} \\ e^{-c/T} & \text{for liquids} \end{cases}. \quad (8.13)$$

It will be shown below that the expression for the stress tensor (8.12) as well as for the temperature dependence of viscosity in the case of gases can be obtained theoretically.

Making use of the expression for σ_{ik} , one can write the equation of motion of an incompressible fluid in the form

$$\rho \left(\frac{\partial v_i}{\partial t} + v_k \frac{\partial v_i}{\partial x_k} \right) = F_i - \frac{\partial p}{\partial x_i} + \eta \frac{\partial}{\partial x_k} \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right)$$

or

$$\rho \left(\frac{\partial v_i}{\partial t} + v_k \frac{\partial v_i}{\partial x_k} \right) = F_i - \frac{\partial p_i}{\partial x_i} + \eta \frac{\partial^2 v_i}{\partial x_k^2}. \quad (8.14)$$

This last equation is called the Navier–Stokes equation.

The Navier–Stokes equation can be rewritten in the form

$$\frac{\partial}{\partial t}(\rho v_i) = - \frac{\partial}{\partial x_k} \left\{ \rho v_i v_k + p \delta_{ik} - \eta \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right) \right\} + F_i$$

or

$$\frac{\partial}{\partial t}(\rho v_i) = - \frac{\partial \Pi_{ik}}{\partial x_k} + F_i,$$

where

$$\Pi_{ik} = \rho v_i v_k + p \delta_{ik} - \eta \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right) = \rho v_i v_k - \sigma_{ik}. \quad (8.15)$$

Integrating both sides of (8.15) over the volume and transforming the integral into a surface integral, we have

$$\frac{\partial}{\partial t} \int \rho v_i dV = - \oint \Pi_{ik} dS_k. \quad (8.16)$$

Clearly, this formula expresses the momentum conservation law: the change of momentum in a given volume of the fluid is equal to the momentum flux expressed in terms of the surface which bounds it. The tensor Π_{ik} represents the momentum flux density tensor for an incompressible fluid.

§9. The energy conservation law and entropy transport in a moving continuous medium

The energy of a moving continuous medium is made up of its internal energy E and kinetic energy $\frac{1}{2}\rho v^2$. In this case we relate the corresponding quantities to unit mass of the fluid and do not take into account the potential energy in an external force field*.

The energy conservation law can be written in integral or differential form

$$\frac{d}{dt} \int (\rho E + \frac{1}{2}\rho v^2) dV = - \oint \mathbf{j}^{(E)} dS, \quad (9.1)$$

* See S.R. de Groot and P. Mazur, *Non-equilibrium thermodynamics* (North-Holland Publishing Company, Amsterdam, 1962).

$$\frac{\partial}{\partial t}(\rho E + \frac{1}{2}\rho v^2) = -\nabla \cdot \mathbf{j}^{(E)}, \quad (9.2)$$

where $\mathbf{j}^{(E)}$ is the energy flux density.

We first of all find the derivative $\partial(\frac{1}{2}\rho v^2)/\partial t$, making use of eq. (8.6). Multiplying this equation by v_i , we have

$$\begin{aligned} \frac{\partial}{\partial t} \frac{1}{2}\rho v^2 &\equiv \frac{\partial}{\partial t} \frac{1}{2}\rho v_i^2 = -v_i \frac{\partial}{\partial x_k} (\rho v_i v_k - \sigma_{ik}) = \\ &= -\frac{\partial}{\partial x_k} (\frac{1}{2}\rho v^2 v_k - \sigma_{ik} v_i) - \sigma_{ik} \frac{\partial v_i}{\partial x_k}. \end{aligned} \quad (9.3)$$

This equation has an obvious interpretation. The first term in the first part represents the kinetic energy flux density. It consists of directly transferred kinetic energy $\frac{1}{2}\rho v^2 \mathbf{v}$ and the flux associated with mechanical work $\sigma_{ik} v_i$ done on the medium. The last term, as can be seen from what follows, involves the energy dissipation due to viscosity.

For an incompressible fluid, one can write, according to (8.12)

$$\begin{aligned} \sigma_{ik} \frac{\partial v_i}{\partial x_k} &= -p \delta_{ik} \frac{\partial v_i}{\partial x_k} + \eta \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right) \frac{\partial v_i}{\partial x_k} = \\ &= -p \frac{\partial v_k}{\partial x_k} + \frac{1}{2}\eta \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right)^2 = \frac{1}{2}\eta \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right)^2. \end{aligned} \quad (9.4)$$

For an incompressible fluid, eq. (9.3) assumes the form

$$\begin{aligned} \frac{\partial}{\partial t} \frac{1}{2}\rho v^2 &= \\ &= -\frac{\partial}{\partial x_k} \left[(\frac{1}{2}\rho v^2 + p) v_k - \eta v_i \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right) \right] - \frac{1}{2}\eta \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right)^2. \end{aligned} \quad (9.5)$$

We see that the change of kinetic energy in a given volume is associated with the energy flow from this volume and with a quantity which does not have the character of flow and which is proportional to the viscosity. It is clear that the latter represents an energy dissipation. Since the dissipation is associated with a decrease in kinetic energy, we always have $\eta > 0$.

Returning to the equation of total energy conservation (9.2), we transform it in such a way as to obtain the law of change of internal energy.

Subtracting (9.3) from (9.2), we find

$$\frac{\partial}{\partial t}(\rho E) = \frac{\partial}{\partial x_k} (\frac{1}{2}\rho v^2 v_k - \sigma_{ik} v_i) + \sigma_{ik} \frac{\partial v_i}{\partial x_k} - \frac{\partial j_k^{(E)}}{\partial x_k}. \quad (9.6)$$

The total energy flux density, $j^{(E)}$, in an isothermal fluid is by definition

made up of the total energy flux transferred by the fluid and of the energy flux due to mechanical work:

$$(j_k^{(E)})_{\text{isotherm}} = (E\rho + \frac{1}{2}\rho v^2)v_k - \sigma_{ik}v_i. \quad (9.7)$$

For $\mathbf{v} = 0$, i.e. in a medium at rest, there is no energy flow.

However, if the fluid is not isothermal, then an energy flow arises in it even if it is at rest. Since the condition of thermodynamic equilibrium is the condition of constancy of temperature, for sufficiently small departures from equilibrium it is natural to set this energy flux equal to

$$\mathbf{j}_T = -\kappa \nabla \cdot T, \quad (9.8)$$

where κ is a kinetic coefficient called the thermal conductivity. The vector \mathbf{j}_T is called the heat flux density. Since the vector \mathbf{j}_T is oriented in the direction of decreasing temperature, we always have $\kappa > 0$.

Of course, the law of transfer (9.7) is empirical and holds only for small departures from thermal equilibrium.

In the general case of a non-isothermal medium, the total energy flux density vector can be written in the form

$$j_k^{(E)} = -\kappa \frac{\partial T}{\partial x_k} + (\rho E + \frac{1}{2}\rho v^2)v_k - \sigma_{ik}v_i. \quad (9.9)$$

Substituting $j_k^{(E)}$ from (9.9) into (9.6), we obtain

$$\frac{\partial \rho E}{\partial t} = -\frac{\partial}{\partial x_k} (\rho E v_k) + \frac{\partial}{\partial x_k} \left(\kappa \frac{\partial T}{\partial x_k} \right) + \sigma_{ik} \frac{\partial v_i}{\partial x_k}. \quad (9.10)$$

Like the kinetic energy, the internal energy of a fluid is not conserved. A conservation law holds only for the sum of these, i.e. for the total energy.

The last formula is conveniently written in another form. That is, one can write

$$\rho \frac{dE}{dt} = \frac{\partial}{\partial t} (\rho E) + \nabla \cdot (\rho E \mathbf{v}). \quad (9.11)$$

Furthermore, we define the heat content in unit volume by the relation

$$\rho \frac{dQ}{dt} = -\nabla \cdot \mathbf{j}_T. \quad (9.12)$$

On separating the term with pressure in the stress tensor σ_{ik} by writing

$$\sigma_{ik} = \sigma'_{ik} - p\delta_{ik}, \quad (9.13)$$

we have (using (9.8), (9.10), (9.11) and (9.12))

$$\rho \frac{dE}{dt} = \rho \frac{dQ}{dt} - p \frac{\partial v_k}{\partial x_k} + \sigma'_{ik} \frac{\partial v_i}{\partial x_k}.$$

Dividing all the equations by ρ , we obtain

$$\frac{dE}{dt} = \frac{dQ}{dt} - \frac{p}{\rho} \frac{\partial v_k}{\partial x_k} + \frac{\sigma'_{ik}}{\rho} \frac{\partial v_i}{\partial x_k}. \quad (9.14)$$

The continuity equation gives

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \rho \frac{\partial v_k}{\partial x_k} = -\rho \frac{\partial v_k}{\partial x_k}. \quad (9.15)$$

Hence, finally,

$$\frac{dE}{dt} = \frac{dQ}{dt} + \frac{p}{\rho^2} \frac{d\rho}{dt} + \frac{\sigma'_{ik}}{\rho} \frac{\partial v_i}{\partial x_k}. \quad (9.16)$$

As we have stressed in §1, we have adopted the assumption of the existence of local equilibrium in a moving medium. Therefore the relation between thermodynamic functions, in particular between internal energy and entropy, is given by the formulae of thermodynamics.

We write the basic thermodynamic equality in the form

$$\frac{dE}{dt} = T \frac{dS}{dt} + \frac{p}{\rho^2} \frac{d\rho}{dt} + \sum \mu_\alpha \frac{dc_\alpha}{dt} \quad (9.17)$$

and eliminate internal energy from (9.17) and (9.16). We then arrive at the equation for entropy balance

$$T \frac{dS}{dt} = \frac{dQ}{dt} + \frac{\sigma'_{ik}}{\rho} \frac{\partial v_i}{\partial x_k} - \sum \mu_\alpha \frac{dc_\alpha}{dt} \quad (9.18)$$

or, making use of (7.8) and (9.12) and writing down the total derivative of the entropy, we find

$$\rho \left(\frac{\partial S}{\partial t} + v_k \frac{\partial S}{\partial x_k} \right) = -\frac{1}{T} \frac{\partial j_k^{(T)}}{\partial x_k} + \frac{\sigma'_{ik}}{T} \frac{\partial v_i}{\partial x_k} + \sum \frac{\partial j_k^{(D,\alpha)}}{\partial x_k} \frac{\mu_\alpha}{T}. \quad (9.19)$$

This equation is conveniently written in the form

$$\begin{aligned} \rho \left(\frac{\partial S}{\partial t} + (\mathbf{v} \cdot \nabla) S \right) = & -\nabla \cdot \left(\frac{\mathbf{j}_T - \sum \mu_\alpha \mathbf{j}^{(D,\alpha)}}{T} \right) + (\mathbf{j}_T \cdot \nabla) \frac{1}{T} - \\ & - \frac{1}{T} \sum_\alpha \mathbf{j}^{(D,\alpha)} \cdot \left((T \cdot \nabla) \frac{\mu_\alpha}{T} \right) + \frac{\sigma'_{ik}}{T} \frac{\partial v_i}{\partial x_k} \end{aligned} \quad (9.20)$$

or, introducing the entropy flow density vector \mathbf{j}_S defined by the equality

$$\mathbf{j}_S = \frac{1}{T} \left(\mathbf{j}_T - \sum \mathbf{j}^{(D, \alpha)} \mu_\alpha \right) \quad (9.21)$$

and denoting by Σ the quantity

$$\Sigma = \mathbf{j}_T \cdot \nabla \frac{1}{T} - \frac{1}{T} \sum_\alpha \mathbf{j}^{(D, \alpha)} \left(T \cdot \nabla \frac{\mu_\alpha}{T} \right) + \frac{\sigma'_{ik}}{T} \frac{\partial v_i}{\partial x_k}, \quad (9.22)$$

we write (9.20) in the form

$$\rho \left(\frac{\partial S}{\partial t} + (\mathbf{v} \cdot \nabla) S \right) = - \nabla \cdot \mathbf{j}_S + \Sigma \quad (9.23)$$

or in the form

$$\frac{\partial \rho S}{\partial t} = - \nabla \cdot (\rho S \mathbf{v} + \mathbf{j}_S) + \Sigma. \quad (9.24)$$

Eq. (9.20) shows that entropy is not conserved: the change of entropy in a given volume per unit time is associated not only with its dispersal with the moving medium via thermal conduction and molecular diffusion, but also with the appearance of the quantity Σ defined by the velocity. This quantity is called the entropy production.

For an incompressible fluid, the last term of the entropy production is, according to (9.4), related to the dissipation due to viscosity. For compressible fluids it has the same meaning but is related to the first as well as second viscosity*.

In a one-component system there is no diffusion flow, $\mathbf{j}_D = 0$, and the corresponding terms of (9.21) and (9.22) reduce to zero. For a single incompressible fluid eq. (9.23) can be written in the form of the equation for temperature.

It is appropriate to recall here that by an incompressible fluid we mean a fluid moving slowly compared with the velocity of sound. For such a motion the density of the medium does not depend explicitly on coordinates and time. Under non-isothermal conditions it can, however, depend on temperature and, consequently, vary in time and space. Nevertheless, the continuity equation for a non-isothermal incompressible fluid is defined by formula (7.4).

The entropy of a fluid can be written as a function of temperature and pressure. We express the derivatives of entropy with respect to coordinates

* See L.D. Landau and E.M. Lifshitz, *Fluid mechanics* (Pergamon Press, London, 1959).

and time in terms of the derivatives of temperature. For this we write the basic equalities

$$\frac{\partial S}{\partial t} = \left(\frac{\partial S}{\partial T} \right)_p \frac{\partial T}{\partial t} = T \left(\frac{\partial S}{\partial T} \right)_p \frac{1}{T} \frac{\partial T}{\partial t} = \frac{C_p}{T} \frac{\partial T}{\partial t},$$

$$\frac{\partial S}{\partial x_i} = \left(\frac{\partial S}{\partial T} \right)_p \frac{\partial T}{\partial x_i} = \frac{C_p}{T} \frac{\partial T}{\partial x_i},$$

$$\mathbf{j}_T = -\kappa \nabla T, \quad \mathbf{j}_D = 0,$$

$$\frac{\sigma'_{ik}}{T} \frac{\partial v_i}{\partial x_k} = \frac{\eta}{T} \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right) \frac{\partial v_i}{\partial x_k} = \frac{\eta}{2T} \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right)^2.$$

Substituting these equalities into (9.20) with $\mathbf{j}_D = 0$, we easily obtain

$$\rho C_p \left(\frac{\partial T}{\partial t} + (\mathbf{v} \cdot \nabla) T \right) = \nabla \cdot (\kappa \nabla T) + \frac{1}{2} \eta \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right)^2. \quad (9.25)$$

This equation represents the equation of thermal conduction for a moving medium.

Further simplifications can often be made in the equation of thermal conduction. Although the kinetic coefficients κ and η are functions of temperature, this dependence can be disregarded for small changes in temperature. In this case the kinetic coefficients of a homogeneous system turn out to be constant at all its points, and can be taken out of the derivative sign. Furthermore, for small velocities of motion the last term is usually (although not always!) small and can be dropped. The equation of thermal conduction then assumes the form

$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = \chi \nabla^2 T, \quad (9.26)$$

where $\chi = \kappa / \rho C_p$ is called the thermal diffusivity.

We have obtained a system of transport equations for moving continuous media. The set of mass transport equations, the equations of motion and the equation of energy (or entropy) transport form a complete system describing the motion of a medium. Indeed, the equations of motion contain five variables (\mathbf{v}, ρ, p) for a single fluid. The entropy equation introduces two more unknown functions; S and T . To determine the seven functions there are three equations of motion, the continuity equation, the entropy equation, the thermodynamic relations $S = S(T, p)$ and the equation of state $p = p(T, \rho)$. The whole set of equations contains three kinetic coefficients — η , ζ and κ .

We shall dwell neither on the boundary conditions with which these equations must be supplemented for their solutions to become single-valued, nor on obtaining these solutions even in the simplest cases. All these problems are discussed in courses in the mechanics of continuous media (or hydrodynamics).

Our purpose has been only to formulate, under certain assumptions, macroscopic (phenomenological) equations for a non-equilibrium continuous medium. It will be shown below that the macroscopic equations themselves, as well as the kinetic coefficients involved in them, can be obtained from kinetic considerations on the basis of molecular concepts, at least for the simplest systems.

§10. The Fokker-Planck equation

We can now go on to the kinetic description of macroscopic systems based on the statistical description.

We shall, first of all, formulate a theory of slow processes which has a very general although somewhat formal character.

Let us consider an arbitrary macroscopic system in a state of incomplete equilibrium. We shall follow changes of its state in the course of the time interval T_0 satisfying the inequality

$$\tau_f \ll T_0 \ll \tau_s, \quad (10.1)$$

where τ_f and τ_s are the relaxation times for the fast and the slow processes in the system.

For incomplete equilibrium states it is possible to define various microscopically determined characteristics of the system. A microscopic state of the system will be characterized by certain parameters, the whole set of which we shall denote by λ . This set of quantities λ will be assumed to run over a continuous sequence of values. The statistical description of the behaviour of the system will be carried out by defining the probability distribution function $\rho(\lambda, t) d\lambda$ which characterizes the probability that at the instant of time t the system is in a state in the interval $\lambda, \lambda + d\lambda$. We shall consider the distribution function to be normalized to unity, $\int \rho(\lambda, t) d\lambda = 1$.

Let the probability for the system to be in state λ_0 at a certain instant of time t be $\rho(\lambda_0, t) d\lambda$. During a time $\Delta t \ll T_0$ the state of the system will change and at the instant of time $t' = t + \Delta t$ it will have the probability $P(\lambda, t + \Delta t) d\lambda$ of being in the state λ .

Our further reasoning will be based on the following assumption. The

probability of getting into the state λ in time Δt is completely determined by defining the state λ_0 and does not depend on how the system got into the state λ_0 , i.e. does not depend on the prehistory of the process. Thus the probability of transition from one arbitrary state into another depends only on this pair of states. In probability theory such states are said to form a Markov chain. Hence our assumption can be formulated briefly as the assumption that transitions between states of the system in time Δt form a Markov chain. In this case the transition probability w can be written in the form $w(\lambda_0, \lambda, \Delta t)$. If the transitions did not have the character of a Markov chain, then w would depend on the state λ_k , in which the system was at $t' < t$ and on the transitions by means of which the system got into the state λ_0 .

We shall consider the probability $w(\lambda_0, \lambda, \Delta t)$ to be normalized to unity

$$\int w(\lambda_0, \lambda, \Delta t) d\lambda = 1. \quad (10.2)$$

Finally, we shall assume that the probability of transition $\lambda_0 \rightarrow \lambda$ decreases rapidly with increasing difference $|\lambda - \lambda_0|$. This means that transitions with a large change of state of the system are unlikely. Transitions for which the state of the system changes relatively little have a high probability. In other words, we assume processes in the system to be slow. Quantitative restrictions imposed upon the properties of $w(\lambda_0, \lambda, \Delta t)$ by this assumption will be given below.

On the basis of the assumptions made one can find a very general equation defining the dependence of the distribution function on the parameters λ and time. For this we note, first of all, that the probability $\rho(\lambda, t + \Delta t) d\lambda$ can be written in the form

$$\rho(\lambda, t + \Delta t) d\lambda = d\lambda \int \rho(\lambda_0, t) w(\lambda_0, \lambda, \Delta t) d\lambda_0, \quad (10.3)$$

where the integration is carried out with respect to all possible values of the variable λ_0 . As a matter of fact, $\rho(\lambda_0, t) w(\lambda_0, \lambda, \Delta t) d\lambda_0 d\lambda$ represents the probability that the system in the state λ_0 will in time Δt go over into the state λ . The total probability that the system at an instant of time $t + \Delta t$ will be in the state λ will be obtained by summing over all possible values of λ_0 .

The integral equation (10.3) is called the Smoluchowski equation. In deriving it we made use only of the assumption that the states of the system form a Markov chain. We now use the assumption that the process is slow. We multiply both sides of (10.3) by an arbitrary function $\varphi(\lambda)/\Delta t$ which is only assumed to be continuous and to have the property $\varphi \rightarrow 0$ as $|\lambda| \rightarrow \infty$, and then integrate over all values of λ . Then we have

$$\frac{1}{\Delta t} \int \rho(\lambda, t + \Delta t) \varphi(\lambda) d\lambda = \frac{1}{\Delta t} \int \rho(\lambda_0, t) d\lambda_0 \int w(\lambda_0, \lambda, \Delta t) \varphi(\lambda) d\lambda. \quad (10.4)$$

Further, we expand the function $\varphi(\lambda)$ in a series in powers of $(\lambda - \lambda_0)$, writing

$$\varphi(\lambda) = \varphi(\lambda_0) + \varphi'(\lambda_0)(\lambda - \lambda_0) + \frac{1}{2}\varphi''(\lambda_0)(\lambda - \lambda_0)^2 + \dots$$

Substituting this expansion into the right-hand side of (10.4), we obtain

$$\begin{aligned} & \int \rho(\lambda_0, t) d\lambda_0 \int [\varphi(\lambda_0) + \varphi'(\lambda_0)(\lambda - \lambda_0) + \\ & \quad + \frac{1}{2}\varphi''(\lambda_0)(\lambda - \lambda_0)^2 + \dots] w(\lambda_0, \lambda, \Delta t) d\lambda = \\ & = \int \rho(\lambda_0, t) \varphi(\lambda_0) d\lambda_0 \int w(\lambda_0, \lambda, \Delta t) d\lambda + \\ & \quad + \int \rho(\lambda_0, t) \varphi'(\lambda_0) d\lambda_0 \int (\lambda - \lambda_0) w(\lambda_0, \lambda, \Delta t) d\lambda + \\ & \quad + \frac{1}{2} \int \rho(\lambda_0, t) \varphi''(\lambda_0) d\lambda_0 \int (\lambda - \lambda_0)^2 w(\lambda_0, \lambda, \Delta t) d\lambda + \dots \end{aligned}$$

Let us consider the inner integrals with respect to the variable λ . The first of these, according to (10.2), is equal to unity.

In view of the rapid decrease in w with increasing $(\lambda - \lambda_0)$, the integrals exist and converge rapidly. It is clear that for such a behaviour of w the successive values of the integrals

$$I_n = \int (\lambda - \lambda_0)^n w d\lambda$$

decrease rapidly with increasing n . We confine ourselves to terms of the second order of small quantities. We then find

$$\begin{aligned} & \int \rho(\lambda_0, t) d\lambda_0 \int \varphi(\lambda) w(\lambda_0, \lambda, \Delta t) d\lambda \simeq \\ & \quad \int \rho(\lambda_0, t) \varphi(\lambda_0) d\lambda_0 + \int \rho(\lambda_0, t) \varphi'(\lambda_0) I_1(\lambda_0, \Delta t) d\lambda_0 + \\ & \quad + \frac{1}{2} \int \rho(\lambda_0, t) \varphi''(\lambda_0) I_2(\lambda_0, \Delta t) d\lambda_0, \end{aligned}$$

Hence

$$\begin{aligned} & \frac{1}{\Delta t} \int \rho(\lambda, t + \Delta t) \varphi(\lambda) d\lambda = \\ & = \frac{1}{\Delta t} \int \rho(\lambda_0, t) \varphi(\lambda_0) d\lambda_0 + \frac{1}{\Delta t} \int \rho(\lambda_0, t) \varphi'(\lambda_0) I_1(\lambda_0, \Delta t) d\lambda_0 + \end{aligned}$$

$$+ \frac{1}{2\Delta t} \int \rho(\lambda_0, t) \varphi''(\lambda_0) I_2(\lambda_0, \Delta t) d\lambda. \quad (10.5)$$

Denoting the integration variable λ_0 of the first integral on the right by λ , transferring the integral to the left-hand side and passing to the limit $\Delta t \rightarrow 0$, we find

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \int \frac{\rho(\lambda, t + \Delta t) - \rho(\lambda, t)}{\Delta t} \varphi(\lambda) d\lambda = \\ = \int \rho(\lambda_0, t) \varphi'(\lambda_0) \left(\lim_{\Delta t \rightarrow 0} \frac{I_1(\lambda_0, \Delta t)}{\Delta t} \right) d\lambda_0 + \\ + \frac{1}{2} \int \rho(\lambda_0, t) \varphi''(\lambda_0) \left(\lim_{\Delta t \rightarrow 0} \frac{I_2(\lambda_0, \Delta t)}{\Delta t} \right) d\lambda_0 \end{aligned}$$

or, introducing the notation

$$\lim_{\Delta t \rightarrow 0} \frac{I_1(\lambda_0, \Delta t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \int \frac{(\lambda - \lambda_0)}{\Delta t} w(\lambda_0, \lambda, \Delta t) d\lambda = a(\lambda_0), \quad (10.6)$$

$$\lim_{\Delta t \rightarrow 0} \frac{I_2(\lambda_0, \Delta t)}{2\Delta t} = \frac{1}{2} \lim_{\Delta t \rightarrow 0} \int \frac{(\lambda - \lambda_0)^2}{\Delta t} w(\lambda_0, \lambda, \Delta t) d\lambda_0 = D(\lambda_0), \quad (10.7)$$

and replacing $\lambda_0 \rightarrow \lambda$ on the right-hand side, we have

$$\int \frac{\partial \rho(\lambda, t)}{\partial t} \varphi(\lambda) d\lambda = \int a(\lambda) \varphi'(\lambda) \rho(\lambda, t) d\lambda + \int D(\lambda) \varphi''(\lambda) \rho(\lambda, t) d\lambda. \quad (10.8)$$

Integrating the two terms on the right-hand side by parts and making use of the properties of the function $\varphi(\lambda)$, we obtain

$$\begin{aligned} \int a(\lambda) \varphi'(\lambda) \rho(\lambda, t) d\lambda &= \varphi(\lambda) \rho(\lambda) a(\lambda) \Big|_{-\infty}^{\infty} - \int \varphi(\lambda) \frac{\partial}{\partial \lambda} (a\rho) d\lambda = \\ &= - \int \varphi(\lambda) \frac{\partial}{\partial \lambda} (a\rho) d\lambda, \\ \int D(\lambda) \varphi''(\lambda) \rho(\lambda, t) d\lambda &= \int \varphi(\lambda) \frac{\partial^2}{\partial \lambda^2} (D\rho) d\lambda. \end{aligned}$$

Substituting these expressions into (10.8) and transferring all the terms to the left-hand side, we find

$$\int \varphi(\lambda) \left[\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial \lambda} (a\rho) - \frac{\partial^2}{\partial \lambda^2} (D\rho) \right] d\lambda = 0.$$

In view of the arbitrariness of $\varphi(\lambda)$, we get finally

$$\frac{\partial \rho}{\partial t} = - \frac{\partial}{\partial \lambda} \left[a \rho - \frac{\partial}{\partial \lambda} D \rho \right] = - \frac{\partial j}{\partial \lambda}. \quad (10.9)$$

Eq. (10.9) is called the Fokker–Planck equation. It defines the dependence of the probability density on the time and on the whole set of parameters λ in the case of an arbitrarily slow process. For brevity of notation, we did not take into account that the quantity λ can represent an arbitrary set of parameters.

In this case, without repeating the calculations, we can write

$$\frac{\partial \rho}{\partial t} = - \frac{\partial}{\partial \lambda_i} \left[a_i \rho - D_{ik} \frac{\partial \rho}{\partial \lambda_k} \right] \quad (10.10)$$

where the summation is carried out over all the values of i and k . The coefficients a_i and D_{ik} are of the form

$$a_i = \lim_{\Delta t \rightarrow 0} \int \frac{(\lambda_i - \lambda_{0i})}{\Delta t} w(\lambda_{0i}, \lambda_i, \Delta t) d\lambda_i,$$

$$D_{ik} = \lim_{\Delta t \rightarrow 0} \int \frac{(\lambda_i - \lambda_{0i})(\lambda_k - \lambda_{0k})}{\Delta t} w(\lambda_{0k}, \lambda_k, \Delta t) d\lambda_i.$$

The Fokker–Planck equation can be given an obvious meaning, if the set of representative points corresponding to a set of identical non-equilibrium systems is considered. In this case the quantity $\rho(\lambda, t)$ can be assumed to be the density of representative points. Then the vector

$$j_i = a_i \rho - \frac{\partial}{\partial \lambda_k} (D_{ik} \rho) \quad (10.11)$$

represents the flux of the representative points in phase space. The Fokker–Planck equation has the mean of the continuity equation in this space. Multiplying (10.10) by λ_i and carrying out the integration over λ_i we obtain

$$\begin{aligned} \int \lambda_i \frac{\partial \rho}{\partial t} d\lambda_i &= \frac{\partial \bar{\lambda}_i}{\partial t} = - \int \dot{\lambda}_i \rho d\lambda_i = - \bar{\dot{\lambda}}_i = \\ &= - \int \lambda_i \frac{\partial}{\partial \lambda_i} \left(a_i \rho - \frac{\partial}{\partial \lambda_i} (D_{ik} \rho) \right) d\lambda_i. \end{aligned}$$

Thus, if we assume that $\rho \rightarrow 0$ for $\lambda \rightarrow \infty$, we have $\bar{\dot{\lambda}}_i = \int a_i \rho d\lambda$. The vector a is the mobility vector. The second term in the vector j does not contribute to the mean velocity. It represents the diffusion current. The tensor D_{ik} is the tensor of the generalized diffusion coefficient.

Frequently the diffusion current j is conveniently rewritten in the form

$$j_i = a'_i - D_{ik} \frac{\partial \rho}{\partial \lambda_k}. \quad (10.12)$$

Then the Fokker-Planck equation transforms to

$$\frac{\partial \rho}{\partial t} = - \frac{\partial}{\partial \lambda_i} \left(a'_i \rho - D_{ik} \frac{\partial \rho}{\partial \lambda_k} \right). \quad (10.13)$$

The formula (10.13) represents the so-called second form of the Fokker-Planck equation. In the particular case when the tensor D_{ik} and the mobility a_i are independent of the variable λ_i , the vector j_i is the direct generalization of the empirical laws (7.10) and (9.8). In this case the Fokker-Planck equation is simplified and turns out to be the usual diffusion equation

$$\frac{\partial \rho}{\partial t} = -a \frac{\partial \rho}{\partial \lambda} + D \frac{\partial^2 \rho}{\partial \lambda^2}.$$

In the stationary state and in the one-dimensional case the Fokker-Planck equation has the form

$$\frac{\partial}{\partial \lambda} \left(a\rho - \frac{\partial}{\partial \lambda} (D\rho) \right) = 0.$$

If the flux of particles tends to zero at infinity, we obtain immediately $a\rho = \partial(D\rho)/\partial\lambda$. Hence

$$\rho(\lambda) = \frac{\text{const}}{D} \exp \left[\int_0^\lambda \frac{a(\lambda')}{D(\lambda')} d\lambda' \right].$$

Let us now consider the case of the Brownian diffusion of a particle taking place in an external field of force. Owing to the action of the external force the probability of transition $w(x_0, x, \Delta t)$, is no longer symmetric with respect to the displacements of the particle in the direction of the force and in the opposite direction. Hence the mean velocity $\bar{v} \neq 0$ and the Fokker-Planck equation can be written in the form

$$\frac{\partial \rho}{\partial t} = - \frac{\partial}{\partial x} (\bar{v}\rho) + D \frac{\partial^2 \rho}{\partial x^2} = - \frac{\partial j}{\partial x}. \quad (10.13')$$

The probability current j is made up of the diffusion flux and the flux due to the action of the force f . As a rule, the mean velocity acquired by the particle can be written in the form $\bar{v} = bf$ (see §56 of Part III). We then obtain the generalized equation of diffusion:

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x} \left(D \frac{\partial \rho}{\partial x} - b f \rho \right). \quad (10.14)$$

For an equilibrium state, eq. (10.14) gives

$$b f \rho - D \frac{\partial \rho}{\partial x} = 0$$

or, in a potential force field,

$$b \frac{\partial U}{\partial x} \rho + D \frac{\partial \rho}{\partial x} = 0$$

or

$$\rho = \text{const } e^{-(bU/D)x}. \quad (10.15)$$

For (10.15) to be the same as the Boltzmann distribution, the Einstein relation must be valid for the connection between the coefficient of diffusion and the mobility $b/D = 1/kT$. Hence (10.14) can be written in the form

$$\frac{\partial \rho}{\partial t} = \frac{D}{kT} \frac{\partial}{\partial x} \left(\frac{\partial U}{\partial x} \rho + kT \frac{\partial \rho}{\partial x} \right) = - \frac{\partial j}{\partial x},$$

where

$$j = -D \frac{\partial \rho}{\partial x} - \frac{D}{kT} \frac{\partial U}{\partial x} \rho. \quad (10.14')$$

For a stationary but non-equilibrium state, $j \neq 0$.

We note that in § 56 of Part III we have already arrived at formula (10.14'), proceeding from an obvious consideration of the process of displacement of a Brownian particle.

Integration of (10.13') gives for the probability distribution

$$\rho(x, t) = (4\pi Dt)^{-\frac{1}{2}} \exp \left\{ - (x + \bar{v}t)^2 / 4Dt \right\} dx,$$

i.e. a Gaussian distribution whose centre moves with the velocity \bar{v} .

As another example, consider the thermal transport of particles through a potential barrier. Let a system of non-interacting particles be in the region of minimum potential energy (potential well). We choose the bottom of the well to be the origin and assume that near the bottom of the well (fig. VI.3) $U = \frac{1}{2} k_1 x^2$. Particles in the potential well are in a state of statistical equilibrium, so that

$$n dx = w(0) e^{-U/kT} dx.$$

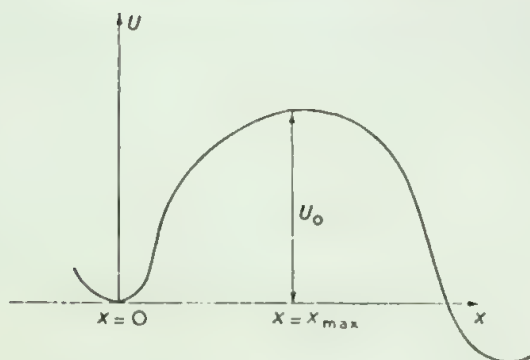


Fig. VI.3 .

Near the point of minimum energy one can write

$$n dx = w(0) e^{-U/kT} dx \approx w(0) e^{-k_1 x^2/2kT} dx ,$$

where $w(0)$ is the probability of finding the particle at the point $x = 0$.

The total number of particles in the well is

$$N = w(0) \int_{-\infty}^{\infty} e^{-k_1 x^2/2kT} dx = w(0) (2\pi kT/k_1)^{1/2} .$$

We have extended the range of integration to the interval $(-\infty, \infty)$ in view of the rapid decrease of the integrand.

In the region $x \sim 1$ let the potential well border upon a potential barrier whose top is at the point $x = 1$. We shall assume that the barrier height U_0 satisfies the inequality

$$U_0 \gg kT . \quad (10.16)$$

Behind the potential barrier the particles again get into a potential well. If the barrier height satisfies inequality (10.16), the number of particles penetrating the barrier is very small. In this case, thermal diffusion through the barrier is such a slow process that it can be considered as stationary.

Making use of (10.14'), we find for the current j

$$j = D \frac{(w e^{U/kT})|_{\infty}^0}{\int e^{U/kT} dx} = \frac{Dw(0)}{\int e^{U/kT} dx} .$$

The potential energy near the maximum can be written in the form

$$U = U_0 - \frac{1}{2}k_2(x - x_{\max})^2.$$

In view of the rapid decrease of the integrand, the integration range can be extended to the entire x -axis. We then find

$$\int e^{U/kT} dx = e^{U_0/kT} \int_{-\infty}^{\infty} e^{-k_2 x^2/2kT} dx = e^{U_0/kT} (2\pi kT/k_2)^{\frac{1}{2}}.$$

Correspondingly

$$j = Dw(0) \left(\frac{k_2}{2\pi kT} \right)^{\frac{1}{2}} e^{-U_0/kT} \quad (10.17)$$

Dividing by the total number of particles at the point $x = 0$, we find the probability that a particle which was initially in one potential well will penetrate through the barrier and get into the other well:

$$P = \frac{j}{N} = \frac{(k_1 k_2)^{\frac{1}{2}}}{2\pi D^{-1}} e^{-U_0/kT}. \quad (10.18)$$

Formula (10.18) is applied for calculating the rates of chemical reactions. In this case, ΔU represents the difference between the energies of the initial and final products.

In what follows we shall frequently have to deal with slow processes and shall see that their behaviour is described by equations of the Fokker–Planck type.

In consequence of its general character, the Fokker–Planck equation gives no detailed information about the behaviour of a system of particles. It has a quasi-macroscopic character and contains unknown coefficients whose value must be determined experimentally or found on the basis of the kinetic description of macroscopic systems.

§11. The basic kinetic equation

The development of an arbitrary quasi-closed subsystem is defined by the equation (2.3) for the density matrix. It is impossible to obtain its exact solution. Hence, instead of an exact equation for the density matrix, use is often made in physical kinetics of the so-called basic kinetic equation which we are now going to derive.

Let us consider a closed system consisting of a large number, N , of inter-

acting particles. We assume that the interaction between the particles was absent for times t , $-\infty < t \leq 0$ and was switched on at the instant of time $t = 0$.

The wave function of the system at $t = 0$ can be represented by the expansion

$$\Psi_0 = \sum c_m(0) \psi_m,$$

where ψ_m is the set of eigenfunctions of certain operators describing the system of particles.

The interaction between the particles leads to a change in the state of the system, and its wave function varies in time $\Psi_0 \rightarrow \Psi(t)$. The wave function $\Psi(t)$ can again be expanded in terms of the functions ψ_m

$$\Psi(t) = \sum c_i(t) \psi_i.$$

We shall assume the interaction to be weak and shall make use of time dependent perturbation theory.

The probability that at an instant of time t the system will be in the i th state is given by the quantity $|c_i(t)|^2$.

According to (55.8) of Part V,

$$c_i(t) = c_i^{(0)} + c_i^{(1)} + \dots, \quad (11.1)$$

where

$$|c_i^{(1)}(t)| = \left| \sum H'_{ik} c_k^{(0)} [1 - \exp [(i/\hbar)(\epsilon_k - \epsilon_i)t] / 2(\epsilon_i - \epsilon_k)] \right|,$$

Here H'_{ik} is the matrix element of the interaction operator \hat{H}' between the wave functions ψ_i .

We now pass from the quantum-mechanical description of the system to the statistical one. For this, as was explained in §2, we replace the values $|c_i(t)|^2$ by their mean values averaged over time, $|\overline{c_i(t)}|^2$. In this case we adopt the random phase hypothesis (2.7). In averaging the square of the sum (11.1), all products $\overline{c_i^{(0)} c_k^{(1)}}$ and $\overline{c_k^{(1)} c_i^{(1)}}$ reduce to zero. We then obtain

$$|\overline{c_i^{(1)}(t)}|^2 = |\overline{c_i^{(0)}}|^2 + \sum |H'_{ik}|^2 |\overline{c_k^{(0)}}|^2 D(t), \quad (11.2)$$

where

$$D(t) = 2 \frac{1 - \cos [(\epsilon_i - \epsilon_k)t/\hbar]}{(\epsilon_i - \epsilon_k)^2}. \quad (11.3)$$

Formula (11.2) defines the probability of transition in time t of particles from the k th state into the i th state. The number of particles in the k th state is proportional to $|c_k|^2$, so that

$$\Delta N_i^+ = \sum |H'_{ik}|^2 N_k D(t).$$

Here ΔN_i^+ is the change in the number of particles in the i th state, associated with it being occupied by particles which were in the k th state.

The balance of particles in the i th state can be written in the form

$$\begin{aligned} \Delta N_i = \Delta N_i^+ - \Delta N_i^- &= \sum |H'_{ik}|^2 N_k(0) D(t) - \\ &- \sum |H'_{ki}|^2 N_i(0) D(t) = \sum |H'_{ik}|^2 (N_k(0) - N_i(0)) D(t). \end{aligned} \quad (11.4)$$

The first term represents the number of particles getting into the i th state in time t , the second term is the number of particles leaving it in the same time. As we have seen in §56 of Part V, for $t \rightarrow \infty$ the factor $D(t)$ is one of the representations of the δ -function:

$$D(t) \rightarrow \frac{\pi}{2\hbar} t \delta(\epsilon_i - \epsilon_k). \quad (11.5)$$

as $t \rightarrow \infty$.

Assuming that the time t is long enough for formula (11.5) to be used, but very short from the macroscopic point of view, we can write

$$\Delta N_i = \frac{\pi}{2\hbar} t \sum |H'_{ik}|^2 (N_k - N_i) \delta(\epsilon_i - \epsilon_k), \quad (11.6)$$

or, since N_i is a microscopically defined quantity, we can replace $\Delta N_i/t$ by the derivative and write finally

$$\frac{dN_i}{dt} = \sum W_{ik} (N_k - N_i), \quad (11.7)$$

where W_{ik} denotes the probability of a transition from the k th into the i th state:

$$W_{ik} = \frac{\pi}{2\hbar} |H'_{ik}|^2 \delta(\epsilon_i - \epsilon_k). \quad (11.8)$$

Eq. (11.7) is called the basic kinetic equation. It plays a very important role in physical kinetics. It can be rewritten in an equivalent form, introducing instead of the number of particles the probability of occupation of a given

state, w_i . We then have

$$\frac{dw_i}{dt} = \sum W_{ik}(w_k - w_i). \quad (11.9)$$

It should be stressed, first of all, what the difference is between the kinetic equation (11.7) and the exact equation for the density matrix (2.4). The kinetic equation contains only the probability w_i but no probability amplitudes. In other words, it contains only the diagonal elements of the density matrix ρ_{nn} .

We have presented its derivation, reproducing to a considerable degree that of formula (56.11) given in §56 of Part V, in order to elucidate the assumptions made and the limits of applicability of the basic kinetic equation.

Let us begin by considering the applicability of formula (11.5). The transition to a δ -function can be made if the interval of values $\Delta\epsilon = \epsilon_i - \epsilon_k$ for all i and k satisfies the condition $\Delta\epsilon \cdot t \sim \hbar$. For a macrosystem $\Delta\epsilon < kT$, so that

$$t > \hbar/kT. \quad (11.10)$$

As we have stressed, the laws of quantum mechanics are reversible and do not change under the substitution $t \rightarrow (-t)$. This is reflected in the symmetry of transition probabilities or in the principle of detailed balance:

$$W_{ik} = W_{ki}. \quad (11.11)$$

We have also seen in quantum mechanics (see §98 of Part V) that the principle of detailed balance is not associated with the application of perturbation theory nor with the concrete form of \hat{H}' in formula (11.9), but is of a general character.

The irreversibility of processes in kinetics arises in averaging the coefficients $|c_i(t)|^2$ over time. This averaging is based on the random phase hypothesis underlying the derivation of the kinetic equation given above.

The random phase hypothesis is used not only for the initial state of the system but also for all other states. In essence, the picture of development described by the kinetic equation (11.7) amounts to the following:

In a certain time t the interaction brings the system from the initial state 1, representing a mixed state, into state 2. Then phases get mixed, and thereupon the system goes over from the mixed state 2 into state 3 and so on. The application of perturbation theory restricts the time t over which formula (11.8) can be used.

Suppose the interaction has the character of collisions and is characterized by a certain time τ . The transition probability must then satisfy the condition $W\tau < 1$. This inequality can be written in the form

$$W < \frac{1}{\tau} \sim \frac{\Delta\epsilon}{\hbar} \sim \frac{kT}{\hbar}. \quad (11.12)$$

It should be noted that, in spite of the wide range of applicability of the basic kinetic equation, the reasoning and assumptions underlying it do not appear to be quite cogent.

The picture of the mixing of phases does not seem to be well substantiated and even seems inconsistent. Indeed, if after a transition the system "forgets" its past, then a unidirectional development in time cannot be understood.

In recent studies there has been success in obtaining a much more cogent derivation of the basic kinetic equation, allowing one not only to give up the phase mixing hypothesis but also to give an answer to the general question of the nature of irreversible processes.

We cannot expound the modern theory and give the derivation of the kinetic equation within the framework of this book. Only the most important concepts of this theory can be presented here*.

We have emphasized that the development in time of macroscopic systems possessing a very large number of degrees of freedom differs essentially from the corresponding processes in microscopic systems possessing a continuous spectrum.

A system undergoing scattering in the continuous spectrum has an infinitely large density of states or, in other words, possesses an infinitely large number of degrees of freedom. However, the character of scattering processes is essentially different from that of processes in macroscopic systems. Scattering processes are strictly reversible in time, whereas processes in non-equilibrium macroscopic systems are always irreversible. Thus the presence of a large number of degrees of freedom in the system does not in itself mean irreversibility. It turns out that the difference between macroscopic systems and microscopic systems with a large number of degrees of freedom is associated with particular features of the form of the Hamiltonian.

Let us consider a closed macroscopic system which at the initial instant of time is in a pure state. The Hamiltonian of the system of particles can be written in the form

$$\hat{H} = \hat{H}_0 + \lambda U, \quad (11.13)$$

where \hat{H}_0 is the Hamiltonian in the absence of any interaction, U represents the interaction energy, and λ is a small parameter.

In scattering processes the interaction energy has singularities at certain

* See, for example, G.V.Chester, The theory of irreversible processes, *Rep. Progress in Physics* V, ch. XXVI, p. 411 (1963).

points of space, whereas in macroscopic systems U is distributed in space throughout the volume of the system. This is the first specific feature of macrosystems.

In the Heisenberg representation

$$\hat{\rho}(t) = e^{(i/\hbar)\hat{H}t} \rho(0) e^{-(i/\hbar)\hat{H}t}, \quad (11.14)$$

so that

$$\langle \hat{L} \rangle = \text{Tr} \{ e^{(i/\hbar)\hat{H}t} \hat{L} e^{-(i/\hbar)\hat{H}t} \rho(0) \}. \quad (11.15)$$

$$e^{-(i/\hbar)\hat{H}t} = e^{-(i/\hbar)\hat{H}_0 t} + (-1)^n \lambda^n \int_0^t dt_1 \int_0^{t_1} dt_2 \dots \int_0^{t_{n-1}} dt_n \times$$

For a Hamiltonian of the form of (11.13) one can write the expansion

$$\times U(t_1) U(t_2) \dots U(t_n), \quad (11.16)$$

where $U(t) = e^{(i/\hbar)\hat{H}_0 t} U(0) e^{-(i/\hbar)\hat{H}_0 t}$. If one substitutes the expansions (11.16) for $e^{(i/\hbar)\hat{H}t}$ and $e^{-(i/\hbar)\hat{H}t}$ into formula (11.15), then the latter will contain terms of the form

$$\lambda^2 \text{Tr} \{ U \hat{L} U \rho(0) \} = \lambda^2 \sum_k \rho(0) \langle \psi_i | U \hat{L} U | \psi_k \rangle \langle \psi_k | U \hat{L} U | \psi_i \rangle, \quad (11.17)$$

where ψ_k is the set of basis functions of the Hamiltonian without interaction \hat{H}_0 .

Terms proportional to higher powers of λ will have an analogous form.

Study of expressions of the type (11.17) shows that with certain restrictions imposed upon the form of the operator U they display a characteristic behaviour: owing to the fact that the interaction energy is distributed throughout the volume of the system of particles, the number of terms in excited states in sum (11.17) for $i = l$ is N times larger than for $i \neq l$.

For a system with a very large number of particles $N \rightarrow \infty$ (N/V being finite), this behaviour of sum (11.17) leads to the appearance of the factor $\delta(\psi_i - \psi_j)$. This allows one to retain the sequence of major terms in the total expression for the mean $\langle L \rangle$ representing an infinite series. Summing these major terms leads automatically to the kinetic equation (11.7). This behaviour of the matrix elements in (11.7) is characteristic only of systems with the property of U mentioned. It does not take place for the corresponding scattering processes in microsystems.

In this derivation it has been shown that no application of the random phase hypothesis to intermediate states is needed. A statistical description of the initial state assumes the random phase distribution only for this state.

Finally, the notion of a weak interaction was defined. That is, weakness of the interaction means that the duration of the process of interaction ('collision') must be small in comparison with the time lapse between two consecutive interactions.

Studies of this kind have confirmed not only the validity and the wide region of applicability of the basic kinetic equation but also established its relation to the time equation for the density matrix. We shall return to this problem later.

§12. Discussion of the basic kinetic equation and some simple examples

We now pass on to the discussion of the basic kinetic equation (11.7). Its simplicity is usually only apparent. Transition probabilities W_{ik} depend on the numbers N_i and N_k . This will be particularly clearly seen in following sections, where it will be shown that a change of state of the particles of an ideal gas arises from collisions, primarily pair collisions. In this last case the probability of transition is proportional to the number of colliding particles in the two states. If one passes from summation to integration, then the kinetic equation (11.7) will transform into a non-linear integro-differential equation. Hence there is only a very limited number of cases in which one can obtain solutions, if only approximate, of the kinetic equation. Some of these will be found below.

We are going to dwell on some simple applications of the basic kinetic equation.

Let us consider, first of all, a simple set of atoms which can be in two states, 1 and 2. Let the energies of these states be ϵ_1 and ϵ_2 . If this system is closed and in a state of statistical equilibrium, then the following equalities hold:

$$-\frac{dN_1}{dt} = W_{12}N_1 - W_{21}N_2 = 0, \quad (12.1)$$

$$-\frac{dN_2}{dt} = W_{21}N_2 - W_{12}N_1 = 0. \quad (12.2)$$

One can apply the principle of detailed balance directly to the closed system and obtain the obvious equality $N_1 = N_2$.

One of the main problems in the theory of irreversible processes is the solution of the question: how can the irreversibility of macroscopic phenomena arise if the behaviour of microscopic particles is strictly reversible. At the

same time the theory of irreversible processes is supposed to answer a lot of more practical questions in such areas as the laws of kinetic behaviour of various macroscopic systems, including the calculation of kinetic coefficients. Within the last ten years great progress has been made in both these directions. The basic kinetic equation was derived in a quite satisfactory way. Further on we shall return to the question of irreversibility in the example of the simplest physical system, i.e. the ideal gas.

Let us now consider a system of atoms in a state of statistical equilibrium with a reservoir. The principle of detailed balance is not directly applicable to such a system. It is applicable only to the closed system 'the set of atoms + the reservoir'.

For the number of atoms in levels 1 and 2 one can write the same expressions (12.1) and (12.2). However, in these $W_{12} \neq W_{21}$. W_{12} now denotes the probability of change of state of the closed system, corresponding to the transition

$$\left. \begin{array}{l} \text{atom in state 1} \\ \text{reservoir in one} \\ \text{of the states } E_0 \end{array} \right\} \rightarrow \left\{ \begin{array}{l} \text{atom in state 2} \\ \text{reservoir in any of the states} \\ \text{with energy } E_0 - (\epsilon_2 - \epsilon_1) . \end{array} \right.$$

Correspondingly, W_{21} represents the probability of the transition

$$\left. \begin{array}{l} \text{atom in state 2} \\ \text{reservoir in any of the states} \\ \text{with energy } E_0 - (\epsilon_2 - \epsilon_1) \end{array} \right\} \rightarrow \left\{ \begin{array}{l} \text{atom in state 1} \\ \text{reservoir in one} \\ \text{of the states with energy } E_0 . \end{array} \right.$$

Since the states of the reservoir in transition $1 \rightarrow 2$, $2 \rightarrow 1$ can be different, it is clear that the principle of detailed balance cannot be applied to these transitions.

We can apply the canonical distribution to the probabilities W_{12} and W_{21} , writing

$$\begin{aligned} \frac{W_{12}}{W_{21}} &= \frac{\Omega(E_0 - (\epsilon_2 - \epsilon_1))}{\Omega(E_0)} = \\ &= e^{\sigma(E_0 - (\epsilon_2 - \epsilon_1)) - \sigma(E_0)} \approx e^{-(\partial \sigma / \partial E)(\epsilon_2 - \epsilon_1)} = e^{-(\epsilon_1 - \epsilon_2)/kT} \end{aligned}$$

Hence we find

$$W_{12} = W_{21} e^{-(\epsilon_2 - \epsilon_1)/kT} \quad (12.3)$$

Then (12.1) and (12.2) lead to the Gibbs distribution:

$$N_2^{(0)} = N_1^{(0)} \frac{W_{12}}{W_{21}} = N_1^{(0)} e^{-(\epsilon_2 - \epsilon_1)/kT} \quad (12.4)$$

Let us now consider the same system in a reservoir, which at the instant of time t is in a non-equilibrium state. Then

$$-\frac{dN_1}{dt} = N_1 W_{12} - W_{21} N_2 = \frac{dN_2}{dt}. \quad (12.5)$$

Suppose that the system is in a state which is near equilibrium. Then we have:

$$N_1 = N_1^{(0)} + N'_1, \quad N'_1 \ll N_1^{(0)}, \quad (12.6)$$

$$N_2 = N_2^{(0)} + N'_2, \quad N'_2 \ll N_2^{(0)}. \quad (12.7)$$

In this case we find

$$-\frac{dN'_1}{dt} = N'_1 W_{12} - W_{21} N'_2 = \frac{dN'_2}{dt}. \quad (12.8)$$

In the kinetic equations (12.8) one can substitute for W_{12} or W_{21} the equilibrium relation (12.3). Then

$$-\frac{d(N'_2 - N'_1)}{dt} = 2(N'_2 W_{21} - N'_1 W_{12}) \approx 2W_{21}(N'_2 - N'_1 e^{-(\epsilon_2 - \epsilon_1)/kT}).$$

For high temperatures $(\epsilon_2 - \epsilon_1)/kT \ll 1$ and, consequently,

$$-\frac{d(N'_2 - N'_1)}{dt} \approx 2W_{21}(N'_2 - N'_1) = \frac{(N'_2 - N'_1)}{\tau},$$

hence

$$N'_2 - N'_1 = (N'_2 - N'_1)_0 e^{-t/\tau}.$$

The number of systems in the non-equilibrium state decreases exponentially with a relaxation time τ equal to

$$\tau = 1/2W_{21}. \quad (12.9)$$

Let us pass to the consideration of a more complex system, where a reservoir contains a system of atoms and radiation.

Atoms in state 1 of energy ϵ_1 can absorb photons of energy $h\nu = \epsilon_2 - \epsilon_1$. The probability of such an absorption is

$$W_{12} = B_{12}\rho(\nu, T),$$

where $\rho(\nu, T)$ is the energy density per unit frequency interval (see § 73 of Part III), and B_{12} is the factor of proportionality. A decrease in the number of atoms in state 1 is given by the expression

$$-\frac{dN_1}{dt} = B_{12}\rho(\nu, T)N_1. \quad (12.10)$$

Atoms in state 2 go over into state 1 as a result of two processes: spontaneous and induced emission. Spontaneous emission takes place in the absence as well as presence of radiation. We shall denote the probability of spontaneous emission for the transition $2 \rightarrow 1$ by A_{21} .

On the basis of the principle of detailed balance an emission process inverse to the absorption process must take place in a system of atoms interacting with radiation. This emission is called induced emission. The probability of induced emission is equal to

$$W_{21} = B_{21}\rho(\nu, T),$$

where

$$B_{21} = B_{12}.$$

Hence the change in the number of atoms in state 2 is given by the formula

$$-\frac{dN_2}{dt} = A_{21}N_2 + B_{21}\rho(\nu, T)N_2. \quad (12.11)$$

If the system (atoms + radiation) is in equilibrium, then

$$\frac{dN_1}{dt} = \frac{dN_2}{dt}$$

and

$$N_1 B_{12}\rho(\nu, T) = A_{21}N_2 + B_{21}\rho(\nu, T)N_2,$$

or

$$\frac{N_2}{N_1} = \frac{B_{12}\rho(\nu, T)}{A_{21} + B_{12}\rho(\nu, T)}. \quad (12.12)$$

On the other hand, for a state of equilibrium N_2/N_1 is expressed by Boltzmann's formula. Hence for the equilibrium density of radiation energy we find

$$\rho(\nu, T) = \frac{A_{21}}{B_{12}} \frac{1}{e^{(\epsilon_2 - \epsilon_1)/kT} - 1} = \frac{A_{21}}{B_{12}} \frac{1}{e^{h\nu/kT} - 1}. \quad (12.13)$$

We thus arrive at the Planck formula.

The ratio A_{21}/B_{12} can be found by passing to the Rayleigh-Jeans formula (i.e. going to a small value of $h\nu/kT \ll 1$).

The derivation of the Planck formula presented above was given by Einstein. Its generalization to the case of atoms with an arbitrary number of levels presents no difficulty.

§ 13. Non-equilibrium systems with a negative temperature. The amplification of electromagnetic waves by such systems

Statistical systems possessing a finite number of levels have certain remarkable features. We stress, first of all, that if all energy levels of the system lie in a limited energy interval, then the argument given in § 17 of Part III, concerning essentially positive values of the temperature, loses validity. Indeed, in § 17 of Part III it was stated that if the temperature θ were less than zero, then the Gibbs distribution could not be made to obey the normalization condition and would be meaningless. However, if the energy runs over only a finite sequence of values, normalization of the Gibbs distribution can be carried out for any value of θ .

Let us consider a system of N atoms with a finite number, n , of energy levels, placed in a reservoir. Although in reality atoms do not have a finite number of levels, for subsequent reasoning it is sufficient that a group of close levels be separated from other levels by a large enough energy interval.

At $T = 0$ all atoms are at the lowest energy level. For $T \gg T_c$, where $T_c = (\epsilon_n - \epsilon_1)/k$, the level distribution becomes uniform. For intermediate values of temperature the number of atoms in a level with given energy, often called the 'level population', is defined by formula (12.4).

Let us now imagine that the atoms are not left to themselves but that energy is supplied to them from outside. We shall consider below one of the ways in which, in practice, energy can be supplied to a system of atoms. If in the system of atoms there is also a mechanism of energy loss, then under certain conditions the system will get into a non-equilibrium but stationary state. The amounts of energy supplied and lost will be equal to each other, and a time-independent energy distribution of atoms will be established in the system. This distribution will be different from the equilibrium one. The level population of the system will be different from (12.4). Namely, under certain conditions, in particular for a sufficiently large value of the energy supplied to the system, the upper level population may become larger than the lower level population.

If we desired to describe the state of our non-equilibrium system in terms of the Gibbs distribution (12.4), we would say that the system has a negative temperature for describing stationary states. Such a terminology is very convenient.

Obviously, the behaviour of the temperature scale is defined by the following reasoning, which we shall, for simplicity, carry out for the example of a system with two levels.

At $T = 0$ all atoms are at the lower level, so that the entropy of the system is equal to zero. When $T \rightarrow \infty$ (actually when $T \gg T_c$) both levels are uniformly occupied, and the entropy is a maximum. It is clear, however, that for $T \rightarrow -\infty$ the same result is obtained: the population of the upper level is equal to that of the lower one. When $|T|$ decreases in the region of negative temperatures the particles progressively go over to the upper level, and for $T \rightarrow 0$ all the particles turn out to be at this level, whereas the lower level is not occupied at all. The entropy of the system is again equal to zero.

Let us now consider quantitatively one of the possible methods of obtaining systems with negative temperature.

Let there be, for example, a plasma in which electrons and atoms (or ions) have different temperatures (i.e. different mean kinetic energies), and let the atoms have three energy levels ϵ_1 , ϵ_2 and ϵ_3 . We shall write the kinetic equation for the number of atoms in these levels.

In collisions with electrons, atoms go over to an upper level with probabilities W_{ik} ($k > i$). Corresponding to such collisions there is transformation of kinetic energy of the electrons into internal energy of the atoms (collisions of the first kind). In addition, on the basis of the principle of detailed balance, there occur collisions of a second kind in which the internal energy of the atoms is transferred to the electrons with probability W_{ki} ($k > i$). Finally, atoms in excited states can emit energy with probability W_{emis} .

Taking into account these processes, we can write for the number of excited particles

$$-\frac{dN_1}{dt} = W_{12}N_1 + W_{13}N_1 - W_{21}N_2 - W_{31}N_3 - W_{\text{emis}}^{(21)}N_2 - W_{\text{emis}}^{(31)}N_3,$$

$$-\frac{dN_2}{dt} = W_{21}N_2 + W_{23}N_2 + W_{\text{emis}}^{(21)}N_2 - W_{12}N_1 - W_{32}N_3 - W_{\text{emis}}^{(32)}N_3,$$

$$-\frac{dN_3}{dt} = W_{31}N_3 + W_{32}N_3 + W_{\text{emis}}^{(31)}N_1 + W_{\text{emis}}^{(32)}N_3 - W_{13}N_1 - W_{23}N_2.$$

In the case where excitation by electrons is compensated by de-excitation

by collisions of the second kind and emission of radiation, a stationary state arises in which all derivatives with respect to time reduce to zero. Assuming the probability of transition from the second to the third level $W_{23} = W_{32}$ to be equal to zero, and also setting $W_{\text{emis}}^{(32)}$ equal to zero (forbidden transition), we obtain from the second equation

$$-W_{21}N_2 + W_{12}N_1 - W_{\text{emis}}^{(21)}N_2 = 0.$$

If, furthermore, the probability of emission $W_{\text{emis}}^{(21)}$ is larger than the probability of collision of the second kind W_{21} , then

$$N_2 \approx \frac{W_{12}}{W_{\text{emis}}^{(21)}} N_1.$$

Under the condition $W_{\text{emis}}^{(31)} \gg W_{31}$, the third equation gives

$$N_3 \approx N_1 \frac{W_{13}}{W_{\text{emis}}^{(31)}} = \frac{W_{13}}{W_{\text{emis}}^{(31)}} \frac{W_{\text{emis}}^{(21)}}{W_{12}} N_2.$$

For the probability ratio W_{13}/W_{12} one can write a formula analogous to (12.3) which involves the temperature of the reservoir with which the energy transfer takes place, in the given case an electron gas with temperature T_{el} :

$$W_{12} = W_{21} e^{-(\epsilon_2 - \epsilon_1)/kT_{\text{el}}},$$

$$W_{13} = W_{31} e^{-(\epsilon_3 - \epsilon_1)/kT_{\text{el}}},$$

$$\frac{W_{12}}{W_{13}} = \frac{W_{21}}{W_{31}} e^{-(\epsilon_3 - \epsilon_2)/kT_{\text{el}}}.$$

Hence, finally,

$$N_2 = \frac{W_{21}}{W_{\text{emis}}^{(21)}} e^{-(\epsilon_2 - \epsilon_1)/kT_{\text{el}}} N_1, \quad (13.1)$$

$$N_3 = \frac{W_{\text{emis}}^{(21)}}{W_{21}} \frac{W_{31}}{W_{\text{emis}}^{(31)}} e^{-(\epsilon_3 - \epsilon_2)/kT_{\text{el}}} N_2. \quad (13.2)$$

The final formula involves the ratios of probabilities of two possible modes of transition from levels 2 and 3, via emission, and via collisions of the second kind. By assumption, the first probability is much higher than the second. Since $\epsilon_2 > \epsilon_1$, it follows from (13.1) that $N_2 \ll N_1$. On the contrary, although

$\epsilon_3 > \epsilon_2$, this formula leads to values $N_3 > N_2$ for a sufficiently large value of the factor in front of the exponential in (13.2). This means that, owing to the fact that the transition from the third to the second level is forbidden, collisions with electrons bring a larger number of atoms to level 3 than to level 2.

Thus level 3 has a negative temperature with respect to level 2. We stress that the procedure considered for obtaining negative temperatures is not the most important and most widely used in practice. However, it makes it possible to elicit the physical conditions necessary for this in the simplest way.

We can now pass on to the properties of systems with negative temperature mentioned above.

If the interaction of a system with radiation is considered for $T < 0$, then it is immediately clear that this interaction differs essentially from that of a system at a temperature $T > 0$. Suppose that the system is acted upon by monochromatic radiation of frequency

$$\nu = \Delta\epsilon/h = (\epsilon_3 - \epsilon_2)/h.$$

This radiation will be absorbed, and the intensity I^- absorbed will be proportional to the number of atoms in state 2 and to the radiation density:

$$I^- = B_{23}N_2\rho.$$

In addition to the absorption, spontaneous and induced emission by atoms in state 3 will take place. The emitted radiation intensity is equal to (see (80.14))

$$I^+ = B_{32}N_3\rho + A_{32}N_3.$$

The difference between the intensities is

$$I^+ - I^- = B_{32}\rho(N_3 - N_2) + A_{32}N_3.$$

If $N_3 - N_2 > 0$, i.e. if level 3 has a negative temperature with respect to level 2, then, on passing through the system, the radiation will not be attenuated by absorption but amplified owing to stimulated emission.

The operation of quantum-mechanical amplifiers and generators (masers with molecular systems or lasers with atomic systems) is based on this effect. They are applied ever more widely in modern radio engineering. Lasers are the most effective generators in the infra-red and optical regions nowadays.

The Kinetic Theory of Gases and Gas-Like Systems

§ 14. Boltzmann's kinetic equation

We have emphasized above that solving the basic kinetic equation is a matter of considerable difficulty. Hence, major practical results have been obtained by considering those physical systems for which the basic kinetic equation can be replaced by a simpler kinetic equation.

The basic kinetic equation defines the numbers N_i , or the state distribution of particles, taking into account all the connections and interactions existing between the particles of the system.

However, it is clear that in a number of cases such a description of macro-systems is too detailed. This is particularly the case for ideal gases.

In the classical approximation, to which we shall restrict ourselves in the following, instead of enumerating the number of particles in a given state we can describe the state of the system by a continuous distribution function. The latter, by virtue of the absence of interaction between the particles, may be separated into a product of distribution functions of individual particles. The definition of the distribution function of an individual particle allows one to describe completely the properties of the ideal gas as a whole.

Let us consider the distribution function for molecules of a non-equilibrium ideal gas. We know that in an ideal gas each molecule can be considered to be a quasi-closed subsystem.

As distinct from the case of an equilibrium gas, the distribution function depends in general on coordinates x, y, z , momentum components p_x, p_y, p_z and time t . In what follows it will be convenient to introduce the following notation. Let dn be the number of molecules whose representative points lie in a phase-space element

$$d\gamma = dx dy dz dp_x dp_y dp_z = d\mathbf{p} dV$$

at the instant of time t . Then

$$dn = f(\mathbf{r}, \mathbf{p}, t) d\gamma,$$

where $f(\mathbf{r}, \mathbf{p}, t)$ is the distribution function sought. In the following we shall make use of the classical approximation. The change in time of the quantity dn (i.e. change in the number of representative points lying in a volume element $d\gamma$) is determined by collisions between the molecules of the gas.

If as a result of a collision between two molecules having momenta \mathbf{p}_1 and \mathbf{p}_2 one of them acquires momentum \mathbf{p} , then its representative point will enter the phase-space element $d\gamma$. If, on the contrary, a molecule having momentum \mathbf{p} collides with another molecule and acquires a new momentum, its representative point leaves the volume $d\gamma$. It is obvious that the larger the volume $d\gamma$, the larger (other things being equal) the number of molecules whose representative points come in and go out of this volume per unit time. The change in the number of particles in an element of phase volume per unit time can be written in the form

$$\frac{df(\mathbf{r}, \mathbf{p}, t) d\gamma}{dt} = (b - a) d\gamma,$$

where $a d\gamma$ is the number of molecules whose representative points go out of the volume element $d\gamma$ as a result of collisions of the type $(\mathbf{p}, \mathbf{p}_1) \rightarrow (\mathbf{p}_2, \mathbf{p}_3)$, and, analogously, $b d\gamma$ is the number of molecules whose representative points come in to the volume element $d\gamma$ as a result of collisions of the type $(\mathbf{p}_2, \mathbf{p}_3) \rightarrow (\mathbf{p}, \mathbf{p}_1)$.

Thus we have

$$\frac{df}{dt} = b - a.$$

Obviously, the total derivative $df(\mathbf{p}, \mathbf{r}, t)/dt$ can be written in the form

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial p_x} \frac{dp_x}{dt} + \frac{\partial f}{\partial p_y} \frac{dp_y}{dt} + \frac{\partial f}{\partial p_z} \frac{dp_z}{dt} + \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial z} \frac{dz}{dt} =$$

$$\frac{\partial f}{\partial t} + \frac{d\mathbf{p}}{dt} \cdot \frac{\partial f}{\partial \mathbf{p}} + \frac{d\mathbf{r}}{dt} \cdot \frac{\partial f}{\partial \mathbf{r}},$$

or

$$\frac{df(\mathbf{r}, \mathbf{p}, t)}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \mathbf{r}} \cdot \frac{\mathbf{p}}{m} + \frac{\partial f}{\partial \mathbf{p}} \cdot \mathbf{F},$$

where $\mathbf{p}/m = \mathbf{v}$ is the velocity of the molecule, and $d\mathbf{p}/dt = \mathbf{F}$ is the force acting upon a molecule having momentum and coordinates lying in the phase-space element $d\gamma$ at the instant of time t .

Finally, we find

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{F} \cdot \frac{\partial f}{\partial \mathbf{p}} = I, \quad (14.1)$$

where $I = b - a$.

The next problem consists of finding the quantity $I = b - a$, which is usually called the collision integral. This integral can be found for a sufficiently rarefied gas in which collisions occur only in pairs. We shall assume that in collisions no transformation of kinetic energy into energy corresponding to other degrees of freedom occurs, i.e. that the collisions between the molecules obey the laws of elastic collisions between rigid spheres.

Momentum and energy conservation laws, which can be written in the form

$$\mathbf{p}_1 + \mathbf{p} = \mathbf{p}_3 + \mathbf{p}_2, \quad (14.2)$$

$$p_1^2 + p^2 = p_3^2 + p_2^2, \quad (14.3)$$

hold for pair-wise elastic collisions between identical particles. The process of elastic collision between molecules can be characterized by the differential cross section for scattering into a solid angle element $d\Omega_1$. This cross section depends on the absolute value of the relative velocity of the colliding particles $v_{\text{rel}} = |\mathbf{v} - \mathbf{v}_1|$ and the scattering angle $\alpha = \alpha(\mathbf{v}, \mathbf{v}_1)$. The differential cross sec-

tion for the scattering of a particle of velocity \mathbf{v} into a solid angle $d\Omega$ can be written in the form

$$d\sigma = \sigma(v_{\text{rel}}, \alpha) d\Omega_1. \quad (14.4)$$

Each of the $f(\mathbf{p}, \mathbf{r}, t) d\mathbf{p} dV$ particles having momentum \mathbf{p} undergoes a certain number of collisions per second with particles which have momentum \mathbf{p}_1 and which are contained in a cylinder of height v_{rel} and base $d\sigma$. This number is just $v_{\text{rel}} d\sigma f(\mathbf{p}_1, \mathbf{r}, t) d\mathbf{p}_1$. Hence the total number of pair collisions undergone per second by the particles confined in the phase-space element $d\gamma$ is equal to

$$v_{\text{rel}} \sigma(v_{\text{rel}}, \alpha) d\Omega_1 f(\mathbf{p}, \mathbf{r}, t) d\mathbf{p} dV f(\mathbf{p}_1, \mathbf{r}, t) d\mathbf{p}_1.$$

To find the quantity $a d\gamma$ in which we are interested, we have to take into account the fact that any collision undergone by a particle with momentum \mathbf{p} leads to a change in \mathbf{p} and to the particle's representative point leaving the volume $d\gamma$. Hence

$$a d\gamma = d\mathbf{p} dV \int \int v_{\text{rel}} \sigma(v_{\text{rel}}, \alpha) f(\mathbf{p}, \mathbf{r}, t) f(\mathbf{p}_1, \mathbf{r}, t) d\mathbf{p}_1 d\Omega_1$$

or

$$a = \int \int v_{\text{rel}} \sigma(v_{\text{rel}}, \alpha) f(\mathbf{p}, \mathbf{r}, t) f(\mathbf{p}_1, \mathbf{r}, t) d\mathbf{p}_1 d\Omega_1. \quad (14.5)$$

The integration is carried out over the entire solid angle Ω_1 and all the values of momentum \mathbf{p}_1 . Similarly one can find the number of collisions as a result of which representative points move into the volume $d\gamma$, i.e. the number of collisions of the type

$$(\mathbf{p}_2, \mathbf{p}_3) \rightarrow (\mathbf{p}, \mathbf{p}_1).$$

Obviously, one can write that the number of collisions of molecules having momenta \mathbf{p}_2 and \mathbf{p}_3 per second is equal to

$$v'_{\text{rel}} \sigma(v'_{\text{rel}}, \alpha) d\Omega_1 f(\mathbf{p}_2, \mathbf{r}, t) f(\mathbf{p}_3, \mathbf{r}, t) dV d\mathbf{p}_2 d\mathbf{p}_3,$$

where $v'_{\text{rel}} = |\mathbf{v}_2 - \mathbf{v}_3|$.

The quantity $b d\gamma$ in which we are also interested is obtained by integrating the last expression with respect to all values of the momenta \mathbf{p}_2 and \mathbf{p}_3 satis-

fying conditions (14.2) and (14.3):

$$b d\gamma = dV \iiint v'_{\text{rel}} \sigma(v'_{\text{rel}}, \alpha) f(\mathbf{p}_2, \mathbf{r}, t) f(\mathbf{p}_3, \mathbf{r}, t) d\mathbf{p}_2 d\mathbf{p}_3 d\Omega_1. \quad (14.6)$$

The elastic collision laws allow v'_{rel} to be expressed in terms of v_{rel} , and $d\mathbf{p}_2 d\mathbf{p}_3$ in terms of $d\mathbf{p} d\mathbf{p}_1$. Namely, the relative velocity after collision $v'_{\text{rel}} = |\mathbf{v}_2 - \mathbf{v}_3|$ is equal to the relative velocity before collision $v_{\text{rel}} = |\mathbf{v} - \mathbf{v}_1|$. To make the transition from integration over $d\mathbf{p}_2 d\mathbf{p}_3$ to $d\mathbf{p} d\mathbf{p}_1$ one can write

$$d\mathbf{p}_2 d\mathbf{p}_3 = \left| \frac{\partial(\mathbf{p}_2, \mathbf{p}_3)}{\partial(\mathbf{p}, \mathbf{p}_1)} \right| d\mathbf{p}_1 d\mathbf{p}.$$

The modulus of the Jacobian of the transformation $\mathbf{p}_2, \mathbf{p}_3 \rightarrow \mathbf{p}, \mathbf{p}_1$ is conveniently calculated by directing the vector \mathbf{p} along the x-axis.

A simple calculation gives

$$\left| \frac{\partial(\mathbf{p}_2, \mathbf{p}_3)}{\partial(\mathbf{p}, \mathbf{p}_1)} \right| = 1,$$

hence

$$d\mathbf{p}_2 d\mathbf{p}_3 = d\mathbf{p}_1 d\mathbf{p}.$$

Substituting this last equality into (14.6), we find

$$b = \int \int v_{\text{rel}} \sigma(v_{\text{rel}}, \alpha) f(\mathbf{p}_2, \mathbf{r}, t) f(\mathbf{p}_3, \mathbf{r}, t) d\mathbf{p}_1 d\Omega_1. \quad (14.7)$$

To carry out the integration, the vectors \mathbf{p}_2 and \mathbf{p}_3 in the functions $f(\mathbf{p}_2, \mathbf{r}, t)$ and $f(\mathbf{p}_3, \mathbf{r}, t)$ must be expressed in terms of \mathbf{p} and \mathbf{p}_1 .

By means of the values b and a which have been found, the collision integral can be written in the symmetric form

$$I = b - a = \int \int v_{\text{rel}} \sigma(v_{\text{rel}}, \alpha) [f(\mathbf{p}_2, \mathbf{r}, t) f(\mathbf{p}_3, \mathbf{r}, t) - f(\mathbf{p}, \mathbf{r}, t) f(\mathbf{p}_1, \mathbf{r}, t)] d\mathbf{p}_1 d\Omega_1. \quad (14.8)$$

Substituting the value of the collision integral into (14.1), we arrive at the

equation

$$\begin{aligned} \frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{F} \cdot \frac{\partial f}{\partial \mathbf{p}} = \iint v_{\text{rel}} \sigma(v_{\text{rel}}, \alpha) [f(\mathbf{p}_2, \mathbf{r}, t) f(\mathbf{p}_3, \mathbf{r}, t) - \\ - f(\mathbf{p}, \mathbf{r}, t) f(\mathbf{p}_1, \mathbf{r}, t)] d\mathbf{p}_1 d\Omega_1. \end{aligned} \quad (14.9)$$

In what follows, instead of $f(\mathbf{p}, \mathbf{r}, t)$, $f(\mathbf{p}_1, \mathbf{r}, t)$, $f(\mathbf{p}_2, \mathbf{r}, t)$ and $f(\mathbf{p}_3, \mathbf{r}, t)$ we shall for brevity write respectively, f , f_1 , f_2 and f_3 , while $\sigma(v_{\text{rel}}, \alpha)$ will be denoted by σ .

In this notation we have

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{F} \cdot \frac{\partial f}{\partial \mathbf{p}} = \iint v_{\text{rel}} \sigma [f_2 f_3 - f f_1] d\mathbf{p}_1 d\Omega_1. \quad (14.10)$$

The integro-differential equation (14.9) (or, in the brief notation, (14.10)) for the distribution function is called Boltzmann's kinetic equation.

The significance of Boltzmann's equation extends far beyond the framework of the physical kinetics of ideal gases. As will be seen from a number of examples to be considered in this and other chapters of this book, a number of physical systems which are in essence very different from an ideal gas, but which formally satisfy the requirements underlying the derivation of Boltzmann's kinetic equation, are described by it.

From the mathematical point of view, Boltzmann's equation represents a non-linear integro-differential equation with partial derivatives. For Boltzmann's equation to assume a concrete meaning, it is necessary to know the dependence of the cross section on the relative velocity and on the scattering angle, as well as the force field acting on the particles of the gas. However, even for the simplest assumptions concerning the form of the function $\sigma(v_{\text{rel}}, \alpha)$ and the character of the force field, it is a matter of considerable difficulty to solve Boltzmann's kinetic equation.

Below we shall give methods of solving Boltzmann's equation. In the meanwhile we note that instead of the momenta of the molecules it is often convenient to use their velocities as variables. Using the variables $(\mathbf{v}, \mathbf{r}, t)$ Boltzmann's equation assumes the form

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{v}} = \int \sigma v_{\text{rel}} [f_2 f_3 - f f_1] d\mathbf{v}_1 d\Omega. \quad (14.11)$$

In what follows we shall change from the variables \mathbf{p} to the variables \mathbf{v} or

from (14.9) to (14.11) without further notice. We shall also not specify the transition from vector to tensor notation in which (14.11) assumes the form

$$\frac{\partial f}{\partial t} + v_i \frac{\partial f}{\partial x_i} + \frac{F_i}{m} \frac{\partial f}{\partial v_i} = \int \sigma v_{\text{rel}} [f_2 f_3 - f f_1] dv_1 d\Omega. \quad (14.12)$$

Boltzmann's equation is sometimes written in a more symmetric representation in which the integration in the collision integral is carried out over all values of the momenta of the colliding particles. Namely, by virtue of the momentum and energy conservation laws, one can write

$$\begin{aligned} b - a &= \int \sigma v_{\text{rel}} [f_2 f_3 - f f_1] d\mathbf{p}_1 d\Omega = \\ &= \int \sigma v_{\text{rel}} [f_2 f_3 - f f_1] \delta(\mathbf{p}_3 + \mathbf{p}_2 - \mathbf{p}_1 - \mathbf{p}) d\mathbf{p}_1 d\mathbf{p}_2 \times \\ &\times \delta(\epsilon_2 + \epsilon_3 - \epsilon - \epsilon_1) d\epsilon_3 d\Omega, \end{aligned} \quad (14.13)$$

where the delta-function containing a vector argument, $\delta(\mathbf{p})$, means

$$\delta(\mathbf{p}) = \delta(p_x) \delta(p_y) \delta(p_z).$$

Instead of integration with respect to the energy ϵ_3 , integration with respect to the momentum p_3 can be introduced, since

$$\begin{aligned} d\epsilon_3 d\Omega \delta(\epsilon_2 + \epsilon_3 - \epsilon_1 - \epsilon) &= 2m \delta(p_2^2 + p_3^2 - p_1^2 - p^2) d\epsilon_3 d\Omega = \\ &= \frac{2p_3}{m} d\Omega dp_3 \delta(p_2^2 + p_3^2 - p_1^2 - p^2) = \delta(p_2^2 + p_3^2 - p_1^2 - p^2) \frac{2dp_3}{p_3}. \end{aligned}$$

Hence (14.8) can be rewritten in the form

$$\begin{aligned} b - a &= \int \sigma v_{\text{rel}} [f_2 f_3 - f f_1] \delta(\mathbf{p}_2 + \mathbf{p}_3 - \mathbf{p}_1 - \mathbf{p}) \times \\ &\times \delta(p_3^2 + p_2^2 - p_1^2 - p^2) \left(\frac{1}{p_3} + \frac{1}{p_2} \right) \frac{dp_1 dp_2 dp_3}{m}. \end{aligned} \quad (14.14)$$

Here instead of $2/p_3$ we have written the symmetric expression $p_2^{-1} + p_3^{-1}$, making use of the complete equivalence of the momenta \mathbf{p}_3 and \mathbf{p}_2 .

§ 15. The basic kinetic equation for correlation functions

The very simple and obvious derivation of Boltzmann's equation given above suffers from a number of shortcomings, both theoretical and practical. Only pair collisions between molecules are taken into account in this derivation for which the pair character of the collisions is essential. It is not possible to see how the derivation can be generalized to the case of triple, quadruple, and more, complex collisions. The entire region of applicability of the general theory is confined to the case of very rarefied gases. In addition, the following important theoretical aspect of the derivation which was presented is not at all clear. On the one hand, it is based on the equations of classical mechanics reversible in time. The motion of particles and their collisions obey strictly defined laws. On the other hand, it follows from Boltzmann's equation that, by consequence of the collisions, molecular chaos is established in the gas, and the entropy of the gas increases monotonically, tending to a certain limit (see § 19). It is not clear at what point the statistical, probabilistic character is introduced into the train of calculations.

Many critics of Boltzmann's work have seen in this a paradoxical and groundless result. It is therefore very important to obtain a more logical and obvious derivation of Boltzmann's equation. We shall present a variant of such a derivation, based on the use of correlation functions*.

In Part III we defined correlation functions ρ_m , for which a system of connected equations is obtained in which correlation functions of lower order are expressed in terms of functions of higher order.

In this case, however, we confined ourselves to correlation functions depending on the coordinates of the particles. We now have to generalize the definition of correlation functions to the case where they also depend on momenta and time.

In order to avoid cumbersome notation, we shall denote the whole set of variables $(p_1, p_2, \dots, p_{3N}; q_1, q_2, \dots, q_{3N}, t)$ by (x_N, t) . The distribution function of a system containing N particles, $\rho(x_N, t)$, must satisfy the general equation

$$\frac{d\rho(x_N, t)}{dt} = 0, \quad (15.1)$$

expressing the law of conservation of the number of representative points in

* See N.N. Bogoliubov, *Problems of dynamic theory in statistical physics*, in 'Studies in statistical mechanics', ed. J. de Boer and G.E. Uhlenbeck, vol. 1 (North-Holland, Amsterdam, 1962).

phase space. We write the total derivative in the form

$$\begin{aligned} \frac{d\rho}{dt} &= \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial p_i} \frac{dp_i}{dt} + \frac{\partial \rho}{\partial q_i} \frac{dq_i}{dt} = \\ &= \frac{\partial \rho}{\partial t} - \frac{\partial \rho}{\partial p_i} \frac{\partial H}{\partial q_i} + \frac{\partial \rho}{\partial q_i} \frac{\partial H}{\partial p_i} = \frac{\partial \rho}{\partial t} - \{H; \rho\} = 0. \end{aligned} \quad (15.2)$$

Eq. (15.2) is called Liouville's equation.

In the case considered, of a system of particles where the Hamiltonian is of the form

$$H = \sum \frac{p_i^2}{2m} + \frac{1}{2} \sum U_{ij}(|q_i - q_j|)$$

and the Poisson bracket $\{H; \rho\}$ is defined by the equality

$$\{H; \rho\} = \sum_i \left(\frac{\partial H}{\partial q_i} \frac{\partial \rho}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial \rho}{\partial q_i} \right), \quad (15.3)$$

the interaction between the particles is assumed to be pair-wise. We introduce the correlation function ρ_s , defining it in the same way as in §47 of Part III:

$$\rho_s(x_1, x_2, \dots, x_s, t) = V^s \int \rho(x_1, x_2, \dots, x_N) dx_{s+1} \dots dx_N. \quad (15.4)$$

In what follows we shall be interested only in the function

$$\rho_1(x_1, t) \equiv \rho_1(\mathbf{r}_1, \mathbf{p}_1, t) = (V/N) f(\mathbf{r}_1, \mathbf{p}_1, t), \quad (15.5)$$

representing to within a factor V the distribution function of the particles of the gas, and in the binary function

$$\rho_{12}(x_1, x_2, t) \equiv \rho_{12}(\mathbf{r}_1, \mathbf{p}_1, \mathbf{r}_2, \mathbf{p}_2, t) = (V^2/N^2) f_{12}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{p}_1, \mathbf{p}_2, t). \quad (15.5')$$

Making use of the definitions (15.2) and (15.3), we find

$$\begin{aligned}
 -\frac{\partial \rho_1}{\partial t} &\equiv (V/N) \frac{\partial f}{\partial t} = -V \int \{H; \rho\} dx_2 \dots dx_N = \\
 &= V \sum_{i=1}^N \left(\frac{\partial H}{\partial p_i} \frac{\partial \rho}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial \rho}{\partial p_i} \right) dx_2 \dots dx_N = \\
 &= V \int \left(\frac{\partial H}{\partial p_1} \frac{\partial \rho}{\partial q_1} - \frac{\partial H}{\partial q_1} \frac{\partial \rho}{\partial p_1} \right) dx_2 \dots dx_N + \\
 &+ V \sum_{i=2}^N \int \left(\frac{\partial H}{\partial p_i} \frac{\partial \rho}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial \rho}{\partial p_i} \right) dx_2 \dots dx_N = \\
 &= V \int \frac{\mathbf{p}_1}{m} \frac{\partial \rho}{\partial \mathbf{r}_1} dx_2 \dots dx_N - \int \frac{\partial \rho}{\partial \mathbf{p}_1} \frac{\partial H}{\partial \mathbf{r}_1} dx_2 \dots dx_N + \\
 &+ V \sum_{i=2}^N \int \left(\frac{\partial H}{\partial p_i} \frac{\partial \rho}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial \rho}{\partial p_i} \right) dx_2 \dots dx_N. \quad (15.6)
 \end{aligned}$$

Evidently, we have

$$\int \frac{\mathbf{p}_1}{m} \cdot \frac{\partial \rho}{\partial \mathbf{r}_1} dx_2 \dots dx_N = \frac{\mathbf{p}_1}{m} \cdot \frac{1}{V} \frac{\partial \rho_1}{\partial \mathbf{r}_1}. \quad (15.7)$$

Further, since all the particles are identical

$$\begin{aligned}
 \int \frac{\partial \rho}{\partial \mathbf{p}_1} \cdot \left(\sum_{i=2}^N \frac{\partial U_{1i}}{\partial \mathbf{r}_1} \right) dx_2 \dots dx_N &= (N-1) \int \frac{\partial \rho}{\partial \mathbf{p}_1} \cdot \frac{\partial U_{12}}{\partial \mathbf{r}_1} dx_2 \dots dx_N = \\
 &= \frac{N-1}{V^2} \int \frac{\partial U_{12}}{\partial \mathbf{r}_1} \cdot \frac{\partial \rho_{12}}{\partial \mathbf{p}_1} d\mathbf{r}_2 d\mathbf{p}_2. \quad (15.8)
 \end{aligned}$$

Finally, the terms of the remaining sum can be written in the form

$$\begin{aligned} & \int dx_3 \dots dx_{i-1} dx_{i+1} \dots dx_N \int \left(\frac{\partial H}{\partial p_i} \frac{\partial \rho}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial \rho}{\partial p_i} \right) dx_i = \\ & = \int dx_3 \dots dx_{i-1} dx_{i+1} \dots dx_N \left[\int \frac{p_i}{m} dp_i \int \frac{\partial \rho}{\partial r_i} dr_i - \right. \\ & \quad \left. - \int \frac{\partial H}{\partial r_i} dr_i \int \frac{\partial \rho}{\partial p_i} dp_i \right]. \end{aligned}$$

For all reasonable assumptions of the form of the distribution function it can be considered that

$$\int \frac{\partial \rho}{\partial r_i} dr_i \rightarrow 0; \quad \int \frac{\partial \rho}{\partial p_i} dp_i \rightarrow 0 \quad (15.9)$$

for an indefinite increase in the integration range. Thus all the terms of the third sum in (15.6) are equal to zero and, making use of (15.7) and (15.8) we obtain, instead of (15.6),

$$\frac{\partial \rho_1}{\partial t} + \frac{p_1}{m} \cdot \frac{\partial \rho_1}{\partial r_1} = \frac{N-1}{V} \int \frac{\partial U_{12}}{\partial r_1} \cdot \frac{\partial \rho_{12}}{\partial p_1} dr_2 dp_2. \quad (15.10)$$

Since U_{12} does not depend on momenta, we can write

$$\frac{\partial U_{12}}{\partial r_1} \cdot \frac{\partial \rho_{12}}{\partial p_1} = \frac{\partial U_{12}}{\partial r_1} \cdot \frac{\partial \rho_{12}}{\partial p_1} - \frac{\partial U_{12}}{\partial p_1} \cdot \frac{\partial \rho_{12}}{\partial r_1} = \{U_{12}; \rho_{12}\},$$

so that (15.10) can be written in the form (dropping the index 1)

$$\frac{\partial \rho_1}{\partial t} + \frac{p}{m} \cdot \frac{\partial \rho_1}{\partial r} = \frac{N-1}{V} \int \{U_{12}; \rho_{12}\} dr_2 dp_2. \quad (15.11)$$

Passing to the limit of an infinitely large system ($N \rightarrow \infty$; $V \rightarrow \infty$) having a large but finite specific volume, $v = V/N$, we have

$$\frac{\partial f}{\partial t} + \frac{p}{m} \cdot \frac{\partial f}{\partial r} = \int \{U_{12}; f_{12}\} dr_2 dp_2. \quad (15.12)$$

Eq. (15.12) defines the law of change of the distribution function $f(\mathbf{r}, \mathbf{p}, t)$ relating coordinates, momenta and time for the particles in the gas. We shall call it the basic kinetic equation for the distribution function f .

As in the case of an equilibrium gas, ρ_1 is expressed in terms of ρ_{12} . In turn, reproducing the previous calculations, it is possible to obtain an analogous equation for $\rho_{12}(x_1, x_2, t)$. It is of similar form and contains the ternary function ρ_{123} . However, we shall not write down this equation, since for a rarefied gas the quantity $1/\nu$ is very small and is a small parameter in the equation for ρ_1 . If we are interested in the value of ρ_1 in the first approximation, with respect to the small parameter $1/\nu$, then we have to substitute the value of ρ_{12} in the zero order approximation in expression (15.12). The latter can be found without the kinetic equation for ρ_{12} .

§ 16. The derivation of Boltzmann's equation from the basic kinetic equation

Let us consider a gas sufficiently rarefied that particles undergo only pair collisions.

For simplicity we shall in what follows, assume the gas to be spatially uniform, so that the distribution function ρ_1 depends only on momenta \mathbf{p} and time t , i.e. $\rho_1(t, x_1) = \rho_1(t, \mathbf{p})$. In this case the binary distribution function should not change when the two particles are displaced by a constant vector \mathbf{a} , i.e.

$$\rho_{12}(t, \mathbf{r}_1 + \mathbf{a}, \mathbf{r}_2 + \mathbf{a}, \mathbf{p}_1, \mathbf{p}_2) = \rho_{12}(t, \mathbf{r}_1, \mathbf{r}_2, \mathbf{p}_1, \mathbf{p}_2).$$

This condition is fulfilled only in the case where ρ_{12} is a function of the distance between the particles, $\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$, $\rho_{12} = \rho_{12}(t, \mathbf{r}_{12}, \mathbf{p}_1, \mathbf{p}_2)$. Obviously, we can introduce the following scale of increasing times: (1) the collision time $\tau_c \sim r_0/\bar{v}$, where r_0 is the radius of the interaction sphere, and \bar{v} is the mean velocity; (2) the mean time between two consecutive collisions, $\tau \sim \lambda/\bar{v}$, where λ is the mean free path. As we shall see in what follows, τ represents the relaxation time in microscopic gas volumes. (3) the macroscopic relaxation time or hydrodynamic time, $\tau_{\text{macro}} \sim L/\bar{v}$, where L is a macroscopic length, for example the size of the container in which the gas is placed.

It is obvious that always

$$\tau_c \ll \tau \ll \tau_{\text{macro}}. \quad (16.1)$$

Let us consider the behaviour of the gas in the course of time intervals Δt such that $\tau_c \ll \Delta t \ll \tau$. In this interval only a few molecules of the gas have

time to undergo collisions. Hence, the behaviour in time of the majority of particles in time intervals Δt is described by the function ρ_1 .

Those few particles which, in an interval Δt have time to undergo collisions are imagined to be combined into pairs. It is clear that the number of pairs is small in comparison with the total number of particles, N . The behaviour of the pairs is evidently described by the binary correlation function ρ_{12} . Our problem is to establish the relationship between ρ_1 and ρ_{12} . This will be done by reasoning which, strictly speaking, is valid only for times of the order of Δt .

Let us consider the behaviour of correlation functions in the limiting case $1/\nu \rightarrow 0$.

The equation for the distribution function for an infinitesimal gas density in a spatially uniform gas assumes the form

$$\frac{d\rho_1(\mathbf{p}, t)}{dt} = 0. \quad (16.2)$$

Formula (16.2) has a simple meaning: in this approximation the particles are moving independently of each other.

The solution of eq. (16.2) is conveniently written in the symbolic form

$$\rho_1(\mathbf{p}, t) = S_t^{(1)} \rho_1(\mathbf{p}, t - \tau_c). \quad (16.3)$$

The operator $S_t^{(1)}$ acts upon the function ρ_1 which depends on the variables x_1 , transforming the function taken at the initial instant of time $t - \tau_c$ into the function at the instant of time t . Since without collisions the particles of a uniform gas are moving with constant momentum,

$$\rho_1(\mathbf{p}, t) = \rho_1(\mathbf{p}, t - \tau_c). \quad (16.4)$$

Analogously, one can write in the approximation $1/\nu \rightarrow 0$ an equation for the correlation function ρ_{12} . Clearly, it is of the form

$$d\rho_{12}^{(0)}/dt = 0. \quad (16.5)$$

Correspondingly, the solution of eq. (16.5) can be written in the symbolic form

$$\rho_{12}^{(0)}(t, x_1, x_2) = S_t^{(2)} \rho_{12}(t - \tau_c, x_1, x_2). \quad (16.6)$$

The operator $S_t^{(2)}$ acts upon the function $\rho_{12}(x_1, x_2)$ which depends on the pair of coordinates x_1 and x_2 , transforming it from its value at the initial instant of time $t - \tau_c$, to its value at the instant of time t .

Eq. (16.6) describes the behaviour of a pair of molecules which are not acted upon by other molecules.

It is clear that such a consideration and eq. (16.5) itself have a strict meaning only for time intervals smaller than the time lapse between consecutive collisions, τ . Since, however, this time is very large in comparison with the collision time, we can also use (16.5) approximately for large times, i.e. we shall assume that $t \rightarrow \infty$. This is the basic assumption of the theory we have given, which was developed by N.N. Bogoliubov.

We write (16.6) in the form

$$\begin{aligned} \rho_{12}^{(0)}(t, x_1, x_2) &= S_t^{(2)} \rho_{12}(t - \tau_c, x_1, x_2) = \\ &= S_t^{(2)} [\rho_{12}(t - \tau_c, x_1, x_2) - \rho_1(t - \tau_c, x_1) \rho_1(t - \tau_c, x_2)] + \\ &+ S_t^{(2)} \rho_1(t - \tau_c, x_1) \rho_1(t - \tau_c, x_2). \end{aligned} \quad (16.7)$$

We pass to the limit $t \rightarrow -\infty$, i.e. consider the correlation function long before the collision. It is obvious that for $t \rightarrow -\infty$ the behaviour of a pair of particles is not correlated, and

$$\lim_{t \rightarrow -\infty} S_t^{(2)} \rho_{12}(t - \tau_c, x_1, x_2) = \rho_1(t - \tau_c, x_1) \rho_1(t - \tau_c, x_2). \quad (16.8)$$

Hence the condition of weakening of correlation can be written

$$\lim_{t \rightarrow -\infty} \rho_{12}(t, x_1, x_2) = \rho_1(t - \tau_c, x_1) \rho_1(t - \tau_c, x_2). \quad (16.9)$$

The distribution function ρ_1 of a spatially uniform gas depends only on the momentum. To the limit $t \rightarrow -\infty$ there correspond the limiting values, \mathbf{P}_1 and \mathbf{P}_2 , of the momenta of the particles which at the instant of time $\tilde{t} = \tau_c$ undergo collision, so that

$$\lim_{t \rightarrow -\infty} S_t^{-(2)} \mathbf{p}_1 = \mathbf{P}_1, \quad (16.10)$$

$$\lim_{t \rightarrow -\infty} S_t^{-(2)} \mathbf{p}_2 = \mathbf{P}_2, \quad (16.11)$$

where \mathbf{P}_1 and \mathbf{P}_2 are the limiting values of the momenta of the particles long before collision, when they are outside the interaction sphere.

It is obvious that \mathbf{P}_1 and \mathbf{P}_2 are connected with the momenta \mathbf{p}_1 and \mathbf{p}_2 of the colliding particles by the relation

$$H = \frac{p_1^2 + p_2^2}{2m} + U(\mathbf{r}_1, \mathbf{r}_2) = \frac{P_1^2 + P_2^2}{2m}. \quad (16.12)$$

Using (16.10) and (16.11) and taking into account (16.5), one can write (16.7) approximately in the form

$$\begin{aligned} \rho_{12}^{(0)}(t, x_1, x_2) &\simeq \lim_{t \rightarrow -\infty} S_t^{(2)}[\rho_{12}(t - \tau_c, x_1, x_2) - \\ &\quad - \rho_1(t - \tau_c, x_1)\rho_1(t - \tau_c, x_2)] + \lim_{t \rightarrow -\infty} S_t^{(2)}\rho_1(t - \tau_c, x_1)\rho_1(t - \tau_c, x_2) = \\ &= \rho_1(0, \mathbf{P}_1)\rho_1(0, \mathbf{P}_2) = \rho_1(t, \mathbf{P}_1)\rho_1(t, \mathbf{P}_2). \end{aligned} \quad (16.13)$$

Formula (16.13) allows the binary distribution function ρ_{12} to be expressed in terms of the distribution function ρ_1 . However, ρ_{12} is a function of the momenta \mathbf{p}_1 and \mathbf{p}_2 of the particles colliding at the instant of time t , whereas in formula (16.13) the functions ρ_1 depend on another argument; the limiting values of the momenta \mathbf{P}_1 and \mathbf{P}_2 before collision.

If we now want to obtain the distribution function for a finite but small value of the parameter $1/\nu$, we have to substitute the binary function of the zero order approximation (16.13) into the right-hand side of eq. (15.11). If we take into account the definitions (15.9) and (15.5'), we can write

$$\frac{\partial f(t, \mathbf{P}_1)}{\partial t} = \int \{U_{12}; f(t, \mathbf{P}_1)f(t, \mathbf{P}_2)\} d\mathbf{r}_2 d\mathbf{p}_2. \quad (16.14)$$

We transform the Poisson brackets, making use of the fact that the momenta \mathbf{P}_1 and \mathbf{P}_2 are constant defined vectors.

Because of this,

$$\{H; f(t, \mathbf{P}_1)f(t, \mathbf{P}_2)\} = 0. \quad (16.15)$$

Using (16.15), we find

$$\left\{ \frac{p_1^2 + p_2^2}{2m} + U_{12}; f(t, \mathbf{P}_1)f(t, \mathbf{P}_2) \right\} = 0,$$

Hence it follows that

$$\begin{aligned} \{U_{12}; f(t, \mathbf{P}_1)f(t, \mathbf{P}_2)\} &= - \left\{ \frac{p_1^2 + p_2^2}{2m}; f(t, \mathbf{P}_1)f(t, \mathbf{P}_2) \right\} = \\ &= - \frac{\mathbf{p}_1}{m} \cdot \frac{\partial [f(t, \mathbf{P}_1)f(t, \mathbf{P}_2)]}{\partial (\mathbf{r}_1 - \mathbf{r}_2)} + \frac{\mathbf{p}_2}{m} \cdot \frac{\partial [f(t, \mathbf{P}_1)f(t, \mathbf{P}_2)]}{\partial (\mathbf{r}_1 - \mathbf{r}_2)} = \\ &= - \frac{\mathbf{p}_1 - \mathbf{p}_2}{m} \cdot \frac{\partial}{\partial \mathbf{r}_{12}} [f(t, \mathbf{P}_1)f(t, \mathbf{P}_2)]. \end{aligned}$$

Thus eq. (16.14) can be written in the form

$$\frac{\partial f(t, \mathbf{p})}{\partial t} = - \int \frac{(\mathbf{p}_2 - \mathbf{p}_1)}{m} \cdot \frac{\partial}{\partial \mathbf{r}_{12}} [f(t, \mathbf{P}_1) f(t, \mathbf{P}_2)] d\mathbf{r}_2 d\mathbf{p}_2 = \int I d\mathbf{p}_2, \quad (16.16)$$

where I denotes the integral

$$I = \int \frac{\mathbf{p}_1 - \mathbf{p}_2}{m} \cdot \frac{\partial}{\partial \mathbf{r}_{12}} [f(t, \mathbf{P}_1) f(t, \mathbf{P}_2)] d\mathbf{r}_2.$$

To calculate this integral we introduce a cylindrical system of coordinates $(\mathbf{r}, \varphi, \xi)$, choosing the vector $\mathbf{p}_2 - \mathbf{p}_1$ to be the positive direction of the polar ξ -axis. We then have

$$x_1 = x_2 + r \sin \varphi,$$

$$y_1 = y_2 + r \cos \varphi,$$

$$z_1 = z_2 + \xi,$$

$$(\mathbf{p}_2 - \mathbf{p}_1) = [(p_2 - p_1), 0, 0],$$

$$d\mathbf{r}_{12} = r dr d\varphi d\xi,$$

so that we can write

$$(\mathbf{p}_2 - \mathbf{p}_1) \cdot \frac{\partial}{\partial \mathbf{r}_{12}} [f(t, \mathbf{P}_1) f(t, \mathbf{P}_2)] = \frac{p_2 - p_1}{m} \frac{\partial}{\partial \xi} [f(t, \mathbf{P}_1) f(t, \mathbf{P}_2)].$$

The momenta \mathbf{P}_1 and \mathbf{P}_2 will now be functions of the coordinates r, φ and ξ . We choose as the origin the point $\mathbf{r}_{12} = 0$, i.e. the point at which the collision takes place. For such a choice of axes we have

$$\begin{aligned} I &= \frac{p_2 - p_1}{m} \int_0^\infty r dr \int_0^{2\pi} d\varphi \int_{-\infty}^\infty \frac{\partial}{\partial \xi} [f(t, \mathbf{P}_1) f(t, \mathbf{P}_2)] d\xi = \\ &= 2\pi \frac{(p_2 - p_1)}{m} \int_0^\infty r dr [f(t, \mathbf{P}_1) f(t, \mathbf{P}_2)] \Big|_{\xi \rightarrow -\infty}^{\xi \rightarrow \infty}. \end{aligned}$$

Let us consider the product

$$f(t, \mathbf{P}_1)f(t, \mathbf{P}_2)|^{\xi \rightarrow \infty} = f(t - t'', \mathbf{P}_1)f(t - t'', \mathbf{P}_2)|^{\xi \rightarrow \infty}, \quad (16.17)$$

where $t'' \leq \tau$. The situation $\xi \rightarrow \infty$ corresponds to particles a large distance apart (outside the interaction range). In this case the particles are going away from each other (this is seen most simply from the fact that $(p_2 - p_1)/m = v_{\text{rel}}$ increases with increasing ξ). This means that if the particles at the instant of time t are a large distance apart, then in the past, at a certain instant of time $t - t''$, they were close to each other, interacted and then separated. The momenta \mathbf{P}_1 and \mathbf{P}_2 in (16.17) denote the momenta of the particles before collision, in particular immediately before undergoing collision. We write this in the form

$$f(\mathbf{P}_1, t)f(\mathbf{P}_2, t)|^{\xi \rightarrow \infty} = f(\mathbf{P}_1^*, t)f(\mathbf{P}_2^*, t), \quad (16.18)$$

where \mathbf{P}_1^* and \mathbf{P}_2^* are the momenta immediately before collision.

The product $f(\mathbf{P}_1, t)f(\mathbf{P}_2, t)|^{\xi \rightarrow -\infty}$ has another meaning. The situation $\xi \rightarrow -\infty$ also corresponds to particles which at the instant of time t are a large distance apart, outside the interaction range. However, these particles did not have time to draw together and collide in the time interval from $t - \tau_c$ to t . They move with constant values of momentum, so that

$$f(\mathbf{P}_2, t)f(\mathbf{P}_1, t)|^{\xi \rightarrow -\infty} = f(\mathbf{p}_2, t)f(\mathbf{p}_1, t), \quad (16.19)$$

where \mathbf{p}_1 and \mathbf{p}_2 are their momenta at the instant of time t . Substituting (16.18) and (16.19) into the integral I , we find

$$I = \frac{p_2 - p_1}{m} \int_0^\infty r dr \int d\varphi [f(\mathbf{P}_1^*, t)f(\mathbf{P}_2^*, t) - f(\mathbf{p}_1, t)f(\mathbf{p}_2, t)]. \quad (16.20)$$

Further, substituting (16.20) into (16.16) we find that the kinetic equation assumes the form

$$\frac{\partial f(\mathbf{p}_1, t)}{\partial t} = \iint r dr d\varphi \int v_{\text{rel}} [f(\mathbf{P}_1^*, t)f(\mathbf{P}_2^*, t) - f(\mathbf{p}_1, t)f(\mathbf{p}_2, t)] d\mathbf{p}_2. \quad (16.21)$$

Here we have introduced the relative velocity of the particles $v_{\text{rel}} = |\mathbf{p}_2 - \mathbf{p}_1|/m$. The integrand involves the momenta \mathbf{P}_1^* and \mathbf{P}_2^* before the colli-

sion. In view of the symmetry of two-body problems with respect to time reversal, one can write besides (16.21)

$$\frac{\partial f(\mathbf{p}_1, t)}{\partial t} = \int r \, dr \, d\varphi \int v_{\text{rel}} [f(\mathbf{p}_3, t)f(\mathbf{p}_4, t) - f(\mathbf{p}_2, t)f(\mathbf{p}_1, t)] \, d\mathbf{p}_2, \quad (16.22)$$

where \mathbf{p}_3 and \mathbf{p}_4 are the momenta after the collision. Instead of the impact parameter r and the azimuthal angle, the solid angle and cross section can be introduced by the relation $r \, dr \, d\varphi = \sigma \, d\Omega$. Carrying out this substitution in (16.21), we arrive at Boltzmann's equation

$$\frac{\partial f(\mathbf{p}_1, t)}{\partial t} = \int v_{\text{rel}} \sigma [f(\mathbf{p}_3, t)f(\mathbf{p}_4, t) - f(\mathbf{p}_2, t)f(\mathbf{p}_1, t)] \, d\mathbf{p}_2 \, d\Omega. \quad (16.23)$$

In the case of a non-uniform gas, one obtains in an analogous way a somewhat modified Boltzmann's equation. We have seen that the transition from the reversible equations of mechanics, Liouville's equation (15.2), to Boltzmann's equation involved a statistical stage, formula (16.13). Formula (16.13) is based from the very beginning on the assumption of asymmetry of the process in time. Fixing the gas states at a given instant of time t , we have analyzed the question as to how the gas got into a given state. In this case the statistical distribution at a given instant of time is not related to that in the past. In other words, the system randomly gets into the situation which we call a collision. Before the collision the 'pair' of particles consisted of non-interacting particles. Their meeting at the instant of time t is a random act. Beginning from this instant and up to the instant $t + \tau$, when the particles of the 'pair' undergo a subsequent collision, their motion is of a determinate character. Thus the derivation of Boltzmann's equation is based on the assumption of time asymmetry and the difference between initial and final states. The behaviour of the system at the times $t - \tau_c$ and $t + \tau$ is essentially different.

The derivation of Boltzmann's equation presented above contains in explicit form an expansion in terms of the small parameter $1/v$.

Making use of the method developed for the expansion in terms of the parameter $1/v$, one can obtain the kinetic equation for gases more dense than those described by Boltzmann's equation. However, this derivation is associated with cumbersome calculations and we shall not present it*.

* See the monograph of N.N. Bogoliubov quoted before, and also G.E. Uhlenbeck and G. Ford, *Lectures in statistical mechanics* (American Mathematical Society, Providence, 1963).

§17. The generalized transport equation and the properties of summational (additive) invariants

From Boltzmann's equation it is possible to obtain a number of important general consequences, which are not associated with finding the explicit form of the distribution function.

Suppose that a gas as a whole performs a motion with mean velocity \mathbf{u} . It turns out that by means of Boltzmann's equation one can find the general equation which must be satisfied by the mean value of an arbitrary function of the velocity of relative motion

$$V_i = v_i - u_i, \quad (17.1)$$

i.e. of the function $\psi(V_i - u_i)$, where $i = x, y, z$. The mean value of this function is

$$\bar{\psi}(\mathbf{r}, t) = \frac{\int \psi f d\mathbf{v}}{\int f d\mathbf{v}} = \frac{1}{N} \int \psi(V_i) f d\mathbf{v}, \quad (17.2)$$

where N , the total number of particles per unit volume, is, generally speaking, a function of coordinates and time, since the distribution function f depends on these variables. We set up the derivative

$$\bar{\psi} \frac{\partial N}{\partial t} + N \frac{\partial \bar{\psi}}{\partial t} = \int \psi(V_i) \frac{\partial f}{\partial t} d\mathbf{v}. \quad (17.3)$$

Making use of Boltzmann's equation, one can write (17.3) in the form

$$\begin{aligned} \bar{\psi} \frac{\partial N}{\partial t} + N \frac{\partial \bar{\psi}}{\partial t} = \\ = - \int \psi(V_i) v_k \frac{\partial f}{\partial x_k} d\mathbf{v} - \int \psi(V_i) \frac{F_k}{m} \frac{\partial f}{\partial v_k} d\mathbf{v} + \int \psi(V_i) I d\mathbf{v}. \end{aligned} \quad (17.4)$$

We transform the integrals involved in (17.4). Obviously, we have

$$\begin{aligned}\int \psi v_k \frac{\partial f}{\partial x_k} d\mathbf{v} &= \frac{\partial}{\partial x_k} \int \psi v_k f d\mathbf{v} = \frac{\partial}{\partial x_k} N(\overline{v_k \psi}) , \\ \int \psi \frac{F_k}{m} \frac{\partial f}{\partial v_k} d\mathbf{v} &= \frac{F_k}{m} \psi f \Big|_{|\mathbf{v}| \rightarrow \infty} - \frac{F_k}{m} \int f \frac{\partial \psi}{\partial v_k} d\mathbf{v} = \\ &= -\frac{F_k}{m} \int f \frac{\partial \psi}{\partial v_k} d\mathbf{v} = -\frac{F_k}{m} N\left(\frac{\partial \psi}{\partial v_k}\right) .\end{aligned}$$

It is assumed here that the distribution function f decreases sufficiently rapidly with increasing absolute value of the velocity, that (ψf) tends to zero as $|\mathbf{v}| \rightarrow \infty$.

Thus (17.4) takes the form

$$\bar{\psi} \frac{\partial N}{\partial t} + N \frac{\partial \bar{\psi}}{\partial t} + \frac{\partial}{\partial x_k} N(\overline{v_k \psi}) - N \frac{F_k}{m} \left(\frac{\partial \bar{\psi}}{\partial v_k} \right) = \int \psi I d\mathbf{v} . \quad (17.5)$$

This equation is called Enskog's equation, or the generalized transport equation. It is clear that in its general form, for an arbitrary form of the function ψ , Enskog's equation is no simpler than Boltzmann's equation. If, however, ψ is one of the additive constants of motion, i.e.

$$\psi(V_i^{\text{rel}}) = m , \quad mV_i , \quad \frac{1}{2}m(V)^2 , \quad (17.6)$$

then the generalized transport equation is substantially simplified.

To see this, let us consider an important general property of the integral standing on the right-hand side of the generalized transport equation. We denote it by G :

$$G = \int \psi I d\mathbf{v} = \int \psi \sigma |\mathbf{v} - \mathbf{v}_1| (f_1 f_3 - f f_1) d\mathbf{v} d\mathbf{v}_1 d\Omega .$$

We symmetrize this expression, making use of the fact that under the replacement of arguments

$$\mathbf{v} \rightarrow \mathbf{v}_1 ; \quad \mathbf{v}_1 \rightarrow \mathbf{v}$$

the integral G does not change

$$G_1 = \int \psi_1 \sigma |\mathbf{v}_1 - \mathbf{v}| (f_2 f_3 - f f_1) d\mathbf{v}_1 d\mathbf{v} d\Omega = G.$$

On the other hand, the pair of variables can be exchanged in such a way as to exchange the velocities before and after collision:

$$\mathbf{v}, \mathbf{v}_1 \rightarrow \mathbf{v}_2, \mathbf{v}_3; \quad \mathbf{v}_2, \mathbf{v}_3 \rightarrow \mathbf{v}, \mathbf{v}_1.$$

Then, obviously, we have

$$|\mathbf{v} - \mathbf{v}_1| = |\mathbf{v}_2 - \mathbf{v}_3|$$

and we again obtain

$$G_2 = \int \psi_2 |\mathbf{v}_2 - \mathbf{v}_3| \sigma [f f_1 - f_2 f_3] d\mathbf{v}_2 d\mathbf{v}_3 d\Omega = -G_1,$$

$$G_3 = \int \psi_3 |\mathbf{v}_3 - \mathbf{v}_2| \sigma [f_1 f - f_3 f_2] d\mathbf{v}_3 d\mathbf{v}_2 d\Omega = -G_1,$$

Summing the quantities G_i we can write them in the symmetric form

$$\begin{aligned} \sum G_i &= \frac{1}{4}(G_1 + G_2 + G_3 + G_4) = \\ &= \frac{1}{4} \int [\psi_2 + \psi_3 - \psi_1 - \psi] \sigma v_{\text{rel}} [f_2 f_3 - f f_1] d\mathbf{v}_1 d\mathbf{v} d\Omega. \end{aligned} \quad (17.7)$$

Making use of the symmetric representation (17.7), the generalized transport equation can be rewritten in the form

$$\begin{aligned} \bar{\psi} \frac{\partial N}{\partial t} + N \frac{\partial \bar{\psi}}{\partial t} + N \frac{\partial}{\partial x_k} (\bar{v}_k \bar{\psi}) - N \frac{F_k}{m} \left(\frac{\partial \bar{\psi}}{\partial v_k} \right) = \\ = -\frac{1}{4} \int [\psi_2 + \psi_3 - \psi - \psi_1] \sigma v_{\text{rel}} (f_2 f_3 - f f_1) d\mathbf{v} d\mathbf{v}_1 d\Omega. \end{aligned} \quad (17.8)$$

We now assume that ψ is one of the additive constants of motion.

The following equalities hold in a collision between particles:

$$m_1 + m_2 = \text{const},$$

$$m_1 \mathbf{v}_1 + m_2 \mathbf{v}_2 = \text{const},$$

$$\frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 = \text{const}.$$

For these five additive constants of the motion

$$\psi_2 + \psi_3 = \psi_1 + \psi ,$$

so that

$$\sum G_i = 0 .$$

Thus G is an additive (summational) integral invariant of collisions. For the five additive integrals the generalized transport equation is simplified and assumes the form

$$\bar{\psi} \frac{\partial N}{\partial t} + N \frac{\partial \bar{\psi}}{\partial t} + N \frac{\partial}{\partial x_k} (\bar{\psi} v_k) - N \frac{F_k}{m} \left(\frac{\partial \bar{\psi}}{\partial v_k} \right) = 0 . \quad (17.9)$$

It is clear that eq. (17.9) is substantially simpler than the general equation (17.8), since it does not contain any non-linear terms describing pair collisions. For this reason eq. (17.9) is more general and is actually applicable even when Boltzmann's equation loses its validity.

§ 18. The equations of motion of a continuous medium

By means of eq. (17.9) it is possible to find the macroscopic laws of momentum and energy density conservation. It is obvious that the macroscopic density of a homogeneous substance can be written in the form

$$\rho = mN ; \quad \psi = m .$$

Then (17.9) gives immediately

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_k} \rho u_k = 0 . \quad (18.1)$$

Eq. (18.1) represents the mass conservation law in a moving medium.

One often has to deal with a many-component gas consisting of molecules of masses m_α . In this case it is necessary to introduce the ideas of the density

and velocity of the α th component:

$$\rho^\alpha = N^\alpha m_\alpha = m_\alpha \int f^{(\alpha)} dv, \quad (18.2)$$

$$u_i^\alpha = \int v_i^\alpha f^{(\alpha)} dv, \quad (18.3)$$

where $\alpha = 1, 2, \dots, r$, and the mean mass density and velocity are

$$\rho = \sum \rho^\alpha = \sum m^\alpha \int f^\alpha dv, \quad (18.4)$$

$$u_k = \frac{1}{\rho} \sum \rho^\alpha u_k^\alpha = \frac{1}{\rho} \sum m^\alpha \int v_k^\alpha f^\alpha dv. \quad (18.5)$$

For the quantity $\partial \rho^\alpha / \partial t$ we can write the two expressions:

$$\begin{aligned} \frac{\partial \rho^\alpha}{\partial t} &= m^\alpha \int \frac{\partial f^\alpha}{\partial t} dv = -m^\alpha \int v_k^\alpha \frac{\partial f^\alpha}{\partial x_k} dv = \\ &= -\frac{\partial}{\partial x_k} m^\alpha \int v_k^\alpha f^\alpha dv = -\frac{\partial \rho^\alpha u_k^\alpha}{\partial x_k} \end{aligned} \quad (18.6)$$

or

$$\begin{aligned} \frac{\partial \rho^\alpha}{\partial t} &= -\frac{\partial}{\partial x_k} m^\alpha \int (v_k^\alpha - u_k) f^\alpha dv - \frac{\partial}{\partial x_k} m^\alpha u_k \int f^\alpha dv = \\ &= -\frac{\partial}{\partial x_k} (j_k^\alpha + u_k \rho^\alpha). \end{aligned} \quad (18.7)$$

Here we have introduced the vector

$$j_k^\alpha = m^\alpha \int (v_k^\alpha - u_k) f^\alpha dv, \quad (18.8)$$

called the diffusion flux of the α th component. The vector j_k^α represents the flux of the α th component with respect to the mean mass velocity u_k (see (7.6)). Formula (18.7) represents the mass conservation law written for a many-component mixture.

Summing the expressions (18.8) for different components and taking into

account that

$$\begin{aligned}\sum j_k^\alpha &= \sum m^\alpha \int (v_k^\alpha - u_k) f^\alpha d\mathbf{v} = \\ &= \sum m^\alpha \int v_k^\alpha f^\alpha d\mathbf{v} - u_k \sum m^\alpha \int f^\alpha d\mathbf{v} = 0 ,\end{aligned}$$

we find the total mass conservation law

$$\frac{\partial \rho}{\partial t} = - \frac{\partial}{\partial x_k} \rho u_k . \quad (18.9)$$

We now pass on to the formulation of the momentum conservation law, writing it first for a one-component system.

To obtain the momentum conservation law, one can substitute the value $\psi = mV_i$ into (17.9). However, it is more convenient to reproduce to a certain extent the derivation of the formula, calculating the derivative of the momentum vector per unit volume

$$\rho u_i = Nm \int v_i f d\mathbf{v} . \quad (18.10)$$

Here we shall confine ourselves to the case of a one-component gas. Using Boltzmann's equation, and taking into account the general property of the summational integral invariant G , we obtain

$$\begin{aligned}\frac{\partial}{\partial t}(\rho u_i) &= Nm \int v_i \frac{\partial f}{\partial t} d\mathbf{v} = \\ &= -Nm \int v_i v_k \frac{\partial f}{\partial x_k} d\mathbf{v} - mN \int v_i \frac{F_k}{m} \frac{\partial f}{\partial v_k} d\mathbf{v} = \\ &= -Nm \int (v_i - u_i)(v_k - u_k) \frac{\partial f}{\partial x_k} d\mathbf{v} + mN u_i u_k \int \frac{\partial f}{\partial x_k} d\mathbf{v} - \\ &\quad - Nm \frac{\partial}{\partial x_k} (u_i u_k) + F_k N \int f \frac{\partial v_i}{\partial v_k} d\mathbf{v} .\end{aligned}$$

Here we have formed the differences $(v_i - u_i)$ and $(v_k - u_k)$ and integrated the last term by parts. We transform the first integral to the derivative of the

integral

$$\begin{aligned}
 & \int (v_i - u_i)(v_k - u_k) \frac{\partial f}{\partial x_k} d\mathbf{v} = \\
 & = \int \frac{\partial}{\partial x_k} [(v_i - u_i)(v_k - u_k)f] d\mathbf{v} - \int f \frac{\partial}{\partial x_k} (v_i - u_i)(v_k - u_k) d\mathbf{v} = \\
 & = \frac{\partial}{\partial x_k} \int (v_i - u_i)(v_k - u_k) f d\mathbf{v} + \\
 & + \int f \left[(v_k - u_k) \frac{\partial u_i}{\partial x_k} + (v_i - u_i) \frac{\partial u_k}{\partial x_k} \right] d\mathbf{v} = \\
 & = \frac{\partial}{\partial x_k} \int (v_i - u_i)(v_k - u_k) f d\mathbf{v} + \frac{\partial u_i}{\partial x_k} \int f(v_k - u_k) d\mathbf{v} + \\
 & + \frac{\partial u_k}{\partial x_k} \int f(v_i - u_i) d\mathbf{v} = \frac{\partial}{\partial x_k} \int (v_i - u_i)(v_k - u_k) f d\mathbf{v} .
 \end{aligned}$$

By definition of the mean, the integrals $\int f(v_i - u_i) d\mathbf{v} = 0$. Furthermore, we have

$$\frac{\partial v_i}{\partial v_k} = \delta_{ik} .$$

We introduce the force vector per unit volume

$$f_k = F_k N \quad (18.11)$$

and the tensor σ_{ik} , called the stress tensor

$$\sigma_{ik} = mN \int (v_i - u_i)(v_k - u_k) f d\mathbf{v} = mN \overline{(v_i - u_i)(v_k - u_k)} . \quad (18.12)$$

By definition the tensor σ_{ik} is symmetric, $\sigma_{ik} = \sigma_{ki}$.

Then we finally obtain

$$\frac{\partial}{\partial t} \rho u_i = - \frac{\partial \sigma_{ik}}{\partial x_k} - \frac{\partial}{\partial x_k} \rho u_i u_k + f_i , \quad (18.13)$$

or, making use of (18.9),

$$\rho \left(\frac{\partial u_i}{\partial t} + u_k \frac{\partial u_i}{\partial x_k} \right) = - \frac{\partial \sigma_{ik}}{\partial x_k} + f_i. \quad (18.14)$$

Eq. (18.14) represents the macroscopic equation of motion of a gaseous medium.

Finally, the law of conservation of the macroscopic energy of a gaseous medium can be obtained in an analogous way. The mean energy per unit volume of a gas performing a macroscopic motion with mean velocity u can be written in the form

$$E = \frac{1}{2} mN \int (v_i - u_i)^2 f \, d\mathbf{v}. \quad (18.15)$$

Differentiating with respect to time, we find, analogous to a previous expression

$$\begin{aligned} \frac{\partial E}{\partial t} &= \frac{1}{2} mN \int (v_i - u_i)^2 \frac{\partial f}{\partial t} \, d\mathbf{v} = \\ &= -\frac{1}{2} mN \int (v_i - u_i)^2 v_k \frac{\partial f}{\partial x_k} \, d\mathbf{v} - \frac{1}{2} mN \int (v_i - u_i)^2 \frac{F_k}{m} \frac{\partial f}{\partial v_k} \, d\mathbf{v} = \\ &= -\frac{1}{2} (I_1 + I_2) mN. \end{aligned} \quad (18.16)$$

We transform the two integrals separately. We have

$$\begin{aligned} I_1 &= \int (v_i - u_i)^2 v_k \frac{\partial f}{\partial x_k} \, d\mathbf{v} = \int (v_i - u_i)^2 (v_k - u_k) \frac{\partial f}{\partial x_k} \, d\mathbf{v} + u_k \int (v_i - u_i)^2 \frac{\partial f}{\partial x_k} \, d\mathbf{v} = \\ &= \frac{\partial}{\partial x_k} \int (v_i - u_i)^2 (v_k - u_k) f \, d\mathbf{v} - \int f \frac{\partial}{\partial x_k} [(v_i - u_i)^2 (v_k - u_k)] \, d\mathbf{v} + \\ &+ \frac{\partial}{\partial x_k} u_k \int (v_i - u_i)^2 f \, d\mathbf{v} - \int f \frac{\partial}{\partial x_k} [u_k (v_i - u_i)^2] \, d\mathbf{v} = \\ &= \frac{\partial}{\partial x_k} \int (v_i - u_i)^2 (v_k - u_k) f \, d\mathbf{v} + \frac{\partial}{\partial x_k} u_k \int (v_i - u_i)^2 f \, d\mathbf{v} - \\ &- \int f v_k \frac{\partial}{\partial x_k} (v_i - u_i)^2 \, d\mathbf{v} = \frac{\partial}{\partial x_k} \frac{(2q_k + 2u_k E)}{mN} + 2 \int f (v_i - u_i) v_k \, d\mathbf{v} \frac{\partial u_i}{\partial x_k} = \\ &= \frac{\partial}{\partial x_k} \frac{2q_k + 2u_k E}{mN} + 2 \frac{\partial u_i}{\partial x_k} \frac{\sigma_{ik}}{mN} + 2 \int f (v_i - u_i) v_k \frac{\partial u_k}{\partial x_k} \, d\mathbf{v} = \\ &= \frac{\partial}{\partial x_k} \frac{(2q_k + 2u_k E)}{mN} + 2 \frac{\partial u_i}{\partial x_k} \frac{\sigma_{ik}}{mN}. \end{aligned}$$

Here we have made use of the definition (18.12) and introduced the notation

$$q_k = \frac{1}{2} m N \int (v_i - u_i)^2 (v_k - u_k) f \, dv. \quad (18.17)$$

It is obvious that the vector q_k represents the energy flux density vector. For the integral I_2 we have

$$\begin{aligned} I_2 &= \frac{F_k}{m} \int (v_i - u_i)^2 \frac{\partial f}{\partial v_k} \, dv = \\ &= \frac{F_k}{m} (v_i - u_i)^2 f \Big|_{|v| \rightarrow \infty} - 2 \frac{F_k}{m} \int (v_i - u_i) \delta_{ik} f \, dv = 0. \end{aligned}$$

Finally, we find

$$\frac{\partial E}{\partial t} = - \frac{\partial}{\partial x_k} (q_k + E u_k) - \sigma_{ik} \frac{\partial u_i}{\partial x_k}. \quad (18.18)$$

Relation (18.18) expresses the energy conservation law. The change of energy in a unit volume is associated with the total energy flow from this volume $q_k + E u_k$ and the work done against internal forces $(\sigma_{ik} \partial u_i / \partial x_k)$. If the energy of an ideal gas is expressed in terms of its temperature, then instead of (18.18) one can write

$$c_v \frac{\partial T}{\partial t} = - \frac{\partial}{\partial x_k} (q_k + c_v T u_k) - \sigma_{ik} \frac{\partial u_i}{\partial x_k}. \quad (18.19)$$

The set of equations (18.9), (18.14) and (18.19) represents the system of equations of a gas in the continuous-medium approximation.

Although these equations are derived for an ideal gas, their region of applicability is much larger. They express the general conservation laws for a continuous medium, and in such a general form are applicable not only to rarefied gases but also to liquids. However, for their actual use it is necessary to find the explicit form of the stress tensor σ_{ik} and of the energy flux vector q_k . The latter in its turn requires that the distribution function f be known.

Below it will be shown that in a particular approximation the distribution function and, correspondingly, also the quantities σ_{ik} and q_k can be found by integrating the Boltzmann equation for an ideal gas.

For liquids one has to content oneself with empirical expressions of σ_{ik} and q_k .

By means of analogous calculations, one can also easily obtain the momentum and energy conservation laws for a mixture of ideal gases.

§19. The laws of increase of entropy

In statistical physics we dealt in detail with the law of increase of entropy. In §25 of Part III the principle of increase of entropy was established. It was shown that when the state of a closed system changes the entropy of the final state is always larger than that of the initial state.

However, within the framework of statistical considerations it is impossible to establish how the transition from the initial to the final state is carried out. In kinetics, it turns out to be possible to investigate the character of change of entropy in greater detail and to show that the entropy of an ideal gas increases monotonically with time.

Let us calculate the change of entropy per unit volume of a monatomic ideal gas, using for this the expression (25.6) of Part III

$$s = -k \int f \ln f \, dv.$$

Differentiating s with respect to time and making use of Boltzmann's equation, we have

$$\begin{aligned} \frac{\partial s}{\partial t} &= -\frac{\partial}{\partial t} k \int f \ln f \, dv = -k \int (1 + \ln f) \frac{\partial f}{\partial t} \, dv = \\ &= k \int (1 + \ln f) v_k \frac{\partial f}{\partial x_k} \, dv - k \int (1 + \ln f) I \, dv - \\ &- k \int (1 + \ln f) \frac{F_k}{m} \frac{\partial f}{\partial v_k} \, dv = k \int (1 + \ln f) v_k \frac{\partial f}{\partial x_k} \, dv - \\ &- k \int I (1 + \ln f) \, dv. \end{aligned}$$

It is easily seen that on integration the third term reduces to zero. We also

transform the first term

$$\begin{aligned}
 \int (1 + \ln f) v_k \frac{\partial f}{\partial x_k} d\mathbf{v} &= \\
 &= \frac{\partial}{\partial x_k} \int (1 + \ln f) v_k f d\mathbf{v} - \int f \frac{\partial}{\partial x_k} [(1 + \ln f) v_k] d\mathbf{v} = \\
 &= \frac{\partial}{\partial x_k} \int f \ln f v_k d\mathbf{v} = - \frac{\partial j_k^s}{\partial x_k}.
 \end{aligned}$$

Here $j_k^s = - \int v_k f \ln f d\mathbf{v}$ is the entropy flux vector.

Making use of the symmetry of the collision integral (17.7), we have

$$\begin{aligned}
 \frac{\partial S}{\partial t} + \frac{\partial j_k^s}{\partial x_k} &= -k \int I(1 + \ln f) d\mathbf{v} = \\
 &= -\frac{1}{4}k \int [\ln f + \ln f_1 - \ln f_2 - \ln f_3] (f_2 f_3 - f_1 f) \sigma v_{\text{rel}} d\mathbf{v} d\mathbf{v}_1 d\mathbf{\Omega}. \quad (19.1)
 \end{aligned}$$

Formula (19.1) defines the change of entropy in a given volume element of the gas. This change is associated, on the one hand, with the entropy flow carried by gas particles and, on the other hand, with the molecular collision processes characterized by the right-hand side of (19.1). We integrate (19.1) over the entire volume V of the closed system. We then obtain

$$\frac{\partial S}{\partial t} = \frac{1}{4}k \int \ln \left(\frac{f_2 f_3}{f f_1} \right) (f_2 f_3 - f f_1) \sigma v_{\text{rel}} d\mathbf{v} d\mathbf{v}_1 d\mathbf{\Omega} dV, \quad (19.2)$$

where S is the total entropy of the system $S = \int s dV$.

The integral

$$\int \frac{\partial j_k^{(S)}}{\partial x_k} dV = \oint j_k dS_k = 0,$$

since the flow at the boundary of a closed system is equal to zero.

It is easily seen that the integrand of (19.2) is essentially positive. Indeed,

$$\ln \left(\frac{f_2 f_3}{f f_1} \right) [f_2 f_3 - f f_1] \geq 0. \quad (19.3)$$

If $f_2 f_3 > ff_1$, then the logarithm and the term in brackets have a positive sign; if $f_2 f_3 < ff_1$, then both are negative. When $f_2 f_3 = ff_1$ the integrand reduces to zero. Since the integral of an essentially positive function is also positive, we see that in a closed system

$$\frac{\partial S}{\partial t} \geq 0. \quad (19.4)$$

Thus it is proved that the entropy of a closed system consisting of an ideal monatomic gas increases monotonically in time or is constant. We see that the kinetic theory establishes the character and details of the mechanism of entropy increase, relating it to intermolecular collisions.

§ 20. Equilibrium and local-equilibrium distributions in an ideal gas

When increasing, the entropy tends to a certain limit, so that

$$\frac{\partial S}{\partial t} = 0 \quad \text{as} \quad t \rightarrow \infty. \quad (20.1)$$

In reality, the entropy reaches its maximum value not for $t \rightarrow \infty$, but after the lapse of a certain relaxation time τ .

The calculation of this relaxation time is a very complex problem. We shall find an approximate value for it below.

It is clear that for (20.1) to be fulfilled when $t > \tau$, the equality sign must hold in formula (19.3), i.e.

$$f_2 f_3 = ff_1. \quad (20.2)$$

Formula (20.2) shows that $\ln f$ is an additive constant of the motion. For collisions between two particles there are five and only five additive constants of the motion*: the integrals of mass, momentum and energy.

* Note. See, for example, A. Sommerfeld, *Thermodynamics and statistical mechanics* (Vol. 5 of Lectures on theoretical physics), (Academic Press, New York, 1956).

In the more general case of the three-body problem there are eight additive integrals: those of energy, momentum, angular momentum and mass; see E. T. Whittaker, *A treatise on analytical dynamics* (Cambridge University Press, Cambridge, 1927). Apparently, this last theorem is of a general character.

Hence $\ln f$ must be a linear function of these quantities,

$$\ln f = am + b_i mv_i + \frac{1}{2} cmv_k^2. \quad (20.3)$$

We note that the solution of the functional equation (20.2) presented in statistical physics was based on the assumption that $f \equiv f(v^2)$. The derivation given here is free from this assumption.

The constants a , b_i and c can be expressed in terms of the number of particles N per cm^3 , the mean macroscopic velocity u_i , and the mean energy of the monatomic gas in an equilibrium state. In particular, if the gas as a whole is at rest and $u_i = 0$, then

$$\int f dv = N, \quad (20.4)$$

$$\int v_i f dv = 0, \quad (20.5)$$

$$N \int \frac{1}{2} mv_k^2 f dv = \frac{3}{2} NkT. \quad (20.6)$$

Whence we find Maxwell's equilibrium distribution

$$f^{(M)} = N \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \exp(-mv^2/2kT). \quad (20.7)$$

In an equilibrium gas, the temperature T and density N have constant values throughout the volume of the gas, and no macroscopic motion occurs in it.

If, however, the temperature and density depend on coordinates and time, and the gas is moving with mean velocity u_i , then the collision integral of the Boltzmann equation reduces to zero if we set

$$f^{(0)} = N \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \exp[-m(\mathbf{v} - \mathbf{u})^2/2kT], \quad (20.8)$$

where N and T are functions of coordinates and time. For brevity we put

$$f^{(0)} = \exp(\alpha + \beta_i v_i + \gamma v^2), \quad (20.9)$$

$$\alpha = N \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \exp(-mu^2/2kT), \quad (20.10)$$

$$\beta_i = mu_i/kT, \quad (20.11)$$

$$\gamma = m/2kT. \quad (20.12)$$

The distribution $f^{(0)}$ represents the solution of Boltzmann's equation, if it reduces not only the right-hand side but also the left-hand side of this equation to zero. For this α , β_i and γ must satisfy the conditions given below.

The distribution (20.8) or (20.9) is called the local Maxwell distribution. Substituting (20.9) into the Boltzmann equation and taking into account that its left-hand side must be equal to zero for any value of v_i , we set the coefficients of different powers of v_i separately equal to zero. We then obtain the system of equations

$$\frac{\partial \alpha}{\partial t} + \frac{F_i \beta_i}{m} = 0, \quad (20.13)$$

$$\frac{\partial \beta_i}{\partial t} + 2\gamma \frac{F_i}{m} + \frac{\partial \alpha}{\partial x_i} = 0, \quad (20.14)$$

$$\delta_{ij} \frac{\partial \gamma}{\partial t} + \frac{1}{2} \left(\frac{\partial \beta_i}{\partial x_j} + \frac{\partial \beta_j}{\partial x_i} \right) = 0, \quad (20.15)$$

$$\frac{\partial \gamma}{\partial x_i} = 0. \quad (20.16)$$

Thus equations (20.13)–(20.16) impose a restriction upon the mean velocity of motion u_i consistent with the local Maxwell distribution. Formula (20.6) shows that the gas temperature T must be constant in space. However, the temperature of the gas as a whole can change in time.

Differentiating (20.15) with respect to the coordinates and taking into account (20.16), we have

$$\frac{\partial^2 \beta_i}{\partial x_k \partial x_i} + \frac{\partial^2 \beta_i}{\partial x_i \partial x_k} = 0,$$

Whence it follows that

$$\frac{\partial^2 \beta_i}{\partial x_k \partial x_i} = 0. \quad (20.17)$$

The solution of eq. (20.17) is

$$u_i = a_i(t) + b_{ij}(t) x_j$$

or

$$u_i = 0. \quad (20.18)$$

The coefficients b_{ij} can be found by substituting (20.18) into (20.13). We then obtain

$$b_{ij} = -b_{ji}, \quad b_{ii} = \frac{m}{2kT^2} \frac{\partial T}{\partial t}.$$

Hence, finally,

$$u_i = a_i(t) + b_{ik} e_{lkj} x_l + \frac{m}{2kT^2} \frac{\partial T}{\partial t} x_i$$

or, in vector form,

$$\mathbf{u} = \mathbf{a}(t) + [\boldsymbol{\omega} \times \mathbf{r}] + \frac{m}{2kT^2} \frac{\partial T}{\partial t} \mathbf{r}. \quad (20.19)$$

Thus any translational and rotational (with angular velocity $\boldsymbol{\omega}$) motion of the gas as a whole, and radial motion whose velocity is defined by the last term of (20.19), are consistent with the local Maxwell distribution.

Condition (20.14), which we have not used, contains forces whose form imposes definite bounds upon the coefficients a_i and b_{ij} . If the gas is confined in a motionless container with impermeable walls, then stationary motion of the type (20.19) cannot occur in it. This means that in (20.18) one must put $u_i = 0$. Setting $\beta_i = 0$ in (20.14), we find

$$\frac{\partial \alpha}{\partial x_i} + 2\gamma \frac{F_i}{m} = 0.$$

If the external forces have a potential, then $F_i = -\partial U / \partial x_i$ and for α we have

$$\alpha = -U/kT + \text{const}.$$

Then the normalized equilibrium distribution assumes the form

$$f^{(M-B)} = N \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} e^{-mv^2/2kT} e^{-U/kT}. \quad (20.20)$$

Thus we have arrived at the natural result; the Maxwell-Boltzmann equilibrium distribution (20.6) is established in the gas rather than the Maxwell local distribution (20.8).

The establishment of the Maxwell-Boltzmann equilibrium distribution is associated with the establishment of space and velocity distributions. The relaxation time for a process in velocity space is of the order of magnitude of $\tau \sim \lambda/\bar{v}$, where λ is the mean free path. Let an arbitrary distribution of particles in space, $\varphi(\mathbf{r}, \mathbf{v}, t)$, be defined at the initial instant of time $t = 0$. In a time of the order of τ the velocity distribution of the molecules at each point of space approaches a local Maxwell distribution, so that

$$\varphi(\mathbf{r}, \mathbf{v}, t) \rightarrow f^{(0)}(\mathbf{r}, \mathbf{v}, t).$$

The density of the gas, N , and its temperature, T , do not have time to take on equilibrium values throughout the gas, and the macroscopic motion of its parts (if it took place at the initial instant of time) does not have time to be damped. In order to stress this fact we have written the parameters \mathbf{r} and t in $f^{(0)}$, in addition to the argument \mathbf{v} . The variation in time of the variables N , T and U is described by macroscopic variables and is characterized by the macroscopic time τ_{macro} . The relaxation time τ_{macro} is of the order of L/c , where L is the size of the macroscopic system, and c is the velocity of propagation of perturbations in the gas. As will be shown in §26, this is none other than the velocity of sound. Thus

$$f^{(0)} \rightarrow f^{(M-B)} \quad \text{for} \quad t \approx \tau_{\text{macro}}.$$

It should be stressed that, as distinct from the Maxwell-Boltzmann distribution, the local Maxwell distribution is never exact but only describes approximately the velocity distribution in a limited volume of the gas.

The local Maxwell distribution is an exact solution of the Boltzmann equation only for a velocity as given by formula (20.19) and a density satisfying eq. (20.14). For other values of u , T and $N(r)$, $f^{(0)}$ represents an approximate solution of the Boltzmann equation (see the next section), valid for time intervals τ in the course of which the macroscopic quantities u , T and N do not have time to change and can be considered simply as constants.

§21. The general theory of the solution of Boltzmann's equation

The results obtained in §17–19 were not concerned with finding explicit solutions of Boltzmann's equation. As we have stressed, solving Boltzmann's equation is a matter of considerable difficulty, both of a mathematical and a physical character. These difficulties are associated with the fact that the actual form of the function $\sigma(v^{\text{rel}}, \alpha)$ for molecular collisions is unknown, and hence the very problem of integrating Boltzmann's equation is somewhat indeterminate. In practice one always considers certain simplified models for which the intermolecular interaction can be represented by one or other sufficiently simple laws depending on the distance between the particles.

The simplest and most often considered models are those of particles which are rigid spheres or of particles interacting according to a law of the type $1/r^n$. These interaction laws are rather arbitrary and their choice is determined mainly by the simplicity of the corresponding formulae for the scattering cross section. Use of these models corresponds to the transition from a real gas system to a particular model. However, solving Boltzmann's equation even for a model still has considerable mathematical difficulties associated with it. At present a number of methods have been devised for its solution. We shall dwell only on the most important of them.

The most effective method of finding the general solution of Boltzmann's equation is the method of moments.

Moments are functions of the form

$$M^{(0)} = \int f \, dv, \quad (21.1)$$

$$M_i^{(1)} = \int v_i f \, dv, \quad (21.2)$$

$$M_{ij}^{(2)} = \int v_i v_j f \, dv, \quad (21.3)$$

$$M_{ijk}^{(3)} = \int v_i v_j v_k f \, dv, \quad (21.4)$$

.....

$$M_{i\dots}^{(N)} = \int v_i \dots v_N f \, dv. \quad (21.5)$$

Such important quantities as density, the mean velocity of particle flux, mo-

mentum and energy fluxes are also moments. For example,

$$\rho = NmM^{(0)}, \quad (21.6)$$

$$u_i = \int v_i f d\mathbf{v} = M_i^{(1)}, \quad (21.7)$$

$$j_i = m \int (v_i - u_i) f d\mathbf{v} = m(M_i^{(1)} - u_i M^{(0)}), \quad (21.8)$$

$$\sigma_{ik} = mN(M_{ij}^{(2)} - u_i u_k M^{(0)}), \quad (21.9)$$

$$q_i = \frac{1}{2} mN \int v_i v^2 f d\mathbf{v} = \frac{1}{2} mN M_{ijj}^{(3)}. \quad (21.10)$$

The idea of the method of moments is as follows.

We write the solution of Boltzmann's equation in the form of a series in terms of orthogonal polynomials. For such polynomials it is natural to choose the Hermite—Sonine polynomials (§10 of Part V), which are conveniently written in the form

$$H_{ijkl\dots}^{(m)} = \frac{(-1)^N}{f^{(0)}} \frac{\partial^m f^{(0)}}{\partial v_i \partial v_j \dots}. \quad (21.11)$$

Thus

$$H^{(0)} = 1,$$

$$H_i^{(1)} = \left(\frac{m}{kT}\right)^{\frac{1}{2}} v_i,$$

$$H_{ik}^{(2)} = \left(\frac{m}{kT}\right) (v_i v_k - \delta_{ik}),$$

.....

The Hermite—Sonine polynomials are orthogonal with weight $f^{(0)}$:

$$\int f^{(0)} H^{(m)} H^{(m')} d\mathbf{v} = 0 \quad \text{for } m \neq m'.$$

We write the distribution function in the form of a series

$$f = f^{(0)} (1 + c_i^{(1)} H_i^{(1)} + c_{ik}^{(2)} H_{ik}^{(2)} + \dots) \quad (21.12)$$

with the coefficients $c_{i\ldots}^{(1)}, c_{ik\ldots}^{(N)}$ depending on coordinates and time. A simple calculation shows that the coefficients of this series may be expressed in terms of moments. Making use of the orthogonality of the Hermite polynomials, it is easily shown that

$$c_i^{(1)} = \frac{u_i}{(kT/m)^{1/2}}; \quad c_{ik}^{(2)} = \frac{\sigma_{ik}}{NkT}; \dots \quad (21.13)$$

Multiplying Boltzmann's equation by the Hermite polynomials $H^{(m)}$ and integrating it over velocities, we have (in the absence of external forces):

$$\int H_{ijk\ldots}^{(m)} \left(\frac{\partial f}{\partial t} + v_k \frac{\partial f}{\partial x_k} \right) dv = \int H_{ijk\ldots}^{(m)} I dv. \quad (21.14)$$

We then substitute (21.14) into the expansion (21.12) and, making use of the orthogonality conditions, arrive at the equations for the coefficients $c_{ij\ldots}^{(m)}$. Since the coefficients $c_{ij\ldots}^{(m)}$ can be expressed in terms of moments $M_{ij\ldots}^{(m)}$, this system can be written in the form of an infinite set of differential equations. These equations are of the form

$$\frac{\partial M_{ijk\ldots}^{(m)}}{\partial t} + \sum_k \frac{\partial}{\partial x_k} M_{ijk\ldots}^{(m)} = \int v_i v_j v_k \dots I dv. \quad (21.15)$$

For a particular interaction model, where the intermolecular interaction is approximated by repulsive forces inversely proportional to the fifth power of scattering (the so-called Maxwellian molecules) the infinite set is cut off and reduces to a finite number of equations. However, for more realistic models such a simplification does not arise. Nevertheless, the importance of the method of moments lies in the fact that, in principle, it allows one to find a closed system of equations with respect to macroscopic quantities (moments), which is equivalent to Boltzmann's equation. The exact solution of this system would be equivalent to the exact solution of Boltzmann's equation. For actual calculations one has to approximate the distribution function by a finite number of terms in the expansion (21.12).

Usually one restricts oneself to terms of the third order. In this case it is most convenient to keep only those moments which have a direct meaning. There are the first thirteen moments — $M^{(0)}, M_i^{(1)}, M_{ik}^{(2)}, M_{ikj}^{(3)}$ — in terms of which the density, momentum and energy fluxes are expressed. In this, the so-called thirteen-moment approximation, the distribution function is of the

form

$$f \approx f_0 \left[\left(1 + \frac{\sigma_{ik}}{2N} \frac{v_i v_k}{(kT)^2} - \frac{q_k v_k}{mN} \left(\frac{m}{kT} \right)^2 \right) \left(1 - \frac{mv^2}{5kT} \right) \right] . \quad (21.16)$$

Thus expansion (21.16) involves the coefficients σ_{ik} and q_k which depend, in general, on coordinates and time.

Substituting the approximate value of f according to formula (21.16) into (21.14), one can arrive at a system of equations for these coefficients. A merit of this system is the fact that the unknown coefficients are directly measurable quantities.

To increase the degree of accuracy, one can increase the number of terms of the series which are retained in the expression approximating the distribution function f^* . We cannot dwell on these cumbersome calculations, the more so as they all refer to gas models with arbitrarily defined scattering cross sections $\sigma(v_{\text{rel}}; \alpha)$.

Boltzmann's equation permits essential simplification in two important cases: (1) if we are interested in changes of state of the gas in time intervals $\Delta t > \tau$, (2) if the gas as a whole is in a state close to equilibrium. In the first case it can be assumed that at a time $t < \Delta t$ a velocity distribution close to the local Maxwell distribution $f^{(0)}$ is established at each point of the gas. Then for times $t > \Delta t$ one can try to find a solution of Boltzmann's equation in the form

$$f(\mathbf{r}, \mathbf{v}, t) \approx f^{(0)}(\mathbf{v}, \mathbf{r}, t) [1 + \varphi(\mathbf{r}, \mathbf{v}, t)] , \quad (21.17)$$

where $\varphi f^{(0)}$ is a small change in the function $f^{(0)}$, i.e. $\varphi \ll 1$.

The function φ describes the development of the gas for times which are large in comparison with τ . This method (called the Chapman-Enskog method) allows one to describe the macroscopic behaviour of the gas, for example to calculate the momentum flux or heat flux in the gas.

If the gas is in a state close to an equilibrium state, then one can set

$$f(\mathbf{r}, \mathbf{v}, t) \approx f^{(M)}(\mathbf{v}) [1 + \varphi(\mathbf{r}, \mathbf{v}, t)] , \quad (21.18)$$

where $f^{(M)}(\mathbf{v})$ is an equilibrium Maxwell distribution, and $\varphi \ll 1$. Such an approximation makes sense if small perturbations are applied to a gas which was initially in equilibrium

* M.Kogan, *Dynamika razrezhonnoy gaza (Dynamics of a rarefied gas)* (Nauka, Moscow, 1967).

In this section we shall confine ourselves to an exposition of the general theory. Examples of the application of the general theory will be given below.

In the two cases mentioned Boltzmann's equation is linearized. We shall begin with the first case.

In substituting (21.17) into Boltzmann's equation, one need retain only terms of the first order of small quantities. In this case, the velocity v_i and the forces F_i giving rise to a departure from a local equilibrium distribution are to be considered small quantities of the first order. Hence substituting (21.17) into the left-hand side of Boltzmann's equation gives

$$\frac{\partial \varphi}{\partial t} + v_k \frac{\partial f^{(0)}}{\partial x_k} + \frac{F_k}{m} \frac{\partial f^{(0)}}{\partial v_k}. \quad (21.19)$$

In the collision integral one can write

$$\begin{aligned} [f_2 f_3 - f f_1] &\simeq f_2^{(0)} f_3^{(0)} (1 + \varphi_2 + \varphi_3) - f^{(0)} f_1^{(0)} (1 + \varphi + \varphi_1) \simeq \\ &\simeq f^{(0)} f_1^{(0)} (\varphi_2 + \varphi_3 - \varphi - \varphi_1). \end{aligned} \quad (21.20)$$

Hence, finally, we arrive at the equation

$$\begin{aligned} \frac{\partial \varphi}{\partial t} + v_k \frac{\partial f^{(0)}}{\partial x_k} + \frac{F_k}{m} \frac{\partial f^{(0)}}{\partial v_k} = \\ = f^{(0)} \int \sigma v_{\text{rel}} f_1^{(0)} (\varphi_2 + \varphi_3 - \varphi_1 - \varphi) dv_1 d\Omega. \end{aligned} \quad (21.21)$$

The equation obtained is a linear non-homogeneous integro-differential equation with respect to the function $\varphi(x_k, v_k, t)$.

In §22 we shall carry out the solution of this equation. Here we shall only point out an important property of the non-homogeneous integral equation (21.21).

We write it in the abbreviated operator form

$$\hat{L}\varphi = A, \quad (21.22)$$

where A is a known function, and L is a linear integral operator.

In order to avoid cumbersome formulae, we write $\hat{L}\varphi$ in the symbolic form

$$\hat{L}\varphi = \int w(v, v_1) (\varphi_1 - \varphi) dv_1, \quad (21.23)$$

where

$$w(v, v_1) = w(v_1, v) .$$

Multiplying (21.23) by an arbitrary function $\psi(\mathbf{v})$, and integrating with respect to $d\mathbf{v}$, we find

$$\int \psi(v) \hat{L}\varphi(v_1) d\mathbf{v} = \int \psi(v) w(v, v_1) [\varphi_1(\mathbf{v}_1) - \varphi(\mathbf{v})] d\mathbf{v} d\mathbf{v}_1 . \quad (21.24)$$

Furthermore, we can write

$$\begin{aligned} \int \psi(v_1) \hat{L}\varphi(v) d\mathbf{v}_1 &= \int \psi(\mathbf{v}_1) w(v_1, v) [\varphi(\mathbf{v}) - \varphi_1(\mathbf{v}_1)] d\mathbf{v} d\mathbf{v}_1 = \\ &= - \int \psi(\mathbf{v}) w(v, v_1) [\varphi_1(\mathbf{v}_1) - \varphi(\mathbf{v})] d\mathbf{v} d\mathbf{v}_1 . \end{aligned} \quad (21.25)$$

Comparing (21.24) and (21.25), we rewrite (21.24) in the symmetric form

$$\int \psi \hat{L}\varphi d\mathbf{v} = \frac{1}{2} \int (\varphi - \varphi_1) (\psi - \psi_1) w d\mathbf{v}_1 d\mathbf{v} . \quad (21.26)$$

Exchanging φ and ψ , we find

$$\int \psi \hat{L}\varphi d\mathbf{v} = \int \varphi \hat{L}\psi d\mathbf{v} . \quad (21.27)$$

Thus \hat{L} is a self-adjoint operator. In the particular case where $\psi = \varphi$, (21.26) gives

$$\int \varphi \hat{L}\varphi d\mathbf{v} = \frac{1}{2} \int (\varphi - \varphi_1)^2 w d\mathbf{v} d\mathbf{v}_1 \geq 0 , \quad (21.28)$$

since $w(v, v_1) \geq 0$.

Let us consider two functions; the function φ which satisfies eq. (21.23) and at the same time the equation

$$\int \varphi \hat{L}\varphi d\mathbf{v} = \int \varphi A d\mathbf{v} \quad (21.29)$$

and the function ψ which satisfies only the integral relation

$$\int \psi \hat{L}\psi d\mathbf{v} = \int \psi A \psi d\mathbf{v} , \quad (21.30)$$

but is not the solution of eq. (21.23).

It is then easily shown that the following inequality holds:

$$\int \varphi \hat{L} \varphi \, dv \geq \int \psi \hat{L} \psi \, dv. \quad (21.31)$$

This inequality means that the solution of the linearized integral Boltzmann's equation corresponds to a maximum of the integral $\int \varphi \hat{L} \varphi \, dv$ with respect to all functions satisfying the condition (21.30).

To prove (21.31), we write the positive (by virtue of (21.25)) quantity

$$\begin{aligned} \int (\varphi - \psi) \hat{L} (\varphi - \psi) \, dv &= \int \varphi \hat{L} \varphi \, dv + \int \psi \hat{L} \psi \, dv - \int \varphi \hat{L} \psi \, dv - \\ &\quad - \int \psi \hat{L} \varphi \, dv = \int \varphi \hat{L} \varphi \, dv + \int \psi \hat{L} \psi \, dv - 2 \int \psi \hat{L} \varphi \, dv \geq 0, \end{aligned}$$

where we have made use of (21.27).

Using (21.22) and (21.30), we find

$$\int (\varphi - \psi) \hat{L} (\varphi - \psi) \, dv = \int \varphi \hat{L} \varphi \, dv - \int \varphi \hat{L} \psi \, dv \geq 0$$

Whence follows the inequality (21.31) which was to be proved.

The extreme properties of the solutions of the integral equation (21.21) make it possible to use ordinary variational methods in order to find them. Thus we choose φ in the form of a linear combination of known functions g_i :

$$\varphi = \sum \alpha_i g_i.$$

By choosing the coefficients α_i in such a way that φ has a maximum value, one can find a function sufficiently close to the true solution.

The second case of the linearization of Boltzmann's equation is obtained by substituting (21.18). Since the Maxwell distribution automatically satisfies Boltzmann's equation, we at once obtain

$$\frac{\partial \varphi}{\partial t} + v_k \frac{\partial \varphi}{\partial x_k} + \frac{F_k}{m} \frac{\partial \varphi}{\partial v_k} = \int \sigma v_{\text{rel}} f^M(v_1) [\varphi_2 + \varphi_3 - \varphi - \varphi_1] \, dv_1 \, d\Omega. \quad (21.32)$$

We thus arrive at a homogeneous linear integro-differential equation of the form

$$\frac{\partial \varphi}{\partial x_i} = \hat{L} \varphi. \quad (21.32')$$

Its solution can easily be found, if we know the solution of the homogeneous integral equation

$$\hat{L}\varphi_i^{(0)} = \lambda_i \varphi_i^{(0)}, \quad (21.33)$$

where λ_i are the eigenvalues and $\varphi_i^{(0)}$ the eigenfunctions (orthogonal and normalized) of the operator \hat{L} .

The solution of (21.32) can be written in the form of an expansion in terms of the system of functions

$$\varphi = \sum \alpha_i \varphi_i^{(0)}. \quad (21.34)$$

In this case it is assumed that the spectrum of the functions $\varphi_i^{(0)}$ has a discrete character. An example of an application of this method will be given in §25 and §26.

We note only that the first five eigenvalues of eq. (21.33) can be indicated straight away. Namely, since the functions

$$\varphi = 1, \quad \varphi = \mathbf{v}, \quad \varphi = \frac{1}{2}mv^2$$

reduce the collision operator to zero and do not depend explicitly on x_i , they are the eigenfunctions of eq. (21.33) which correspond to the eigenvalues of the operator \hat{L} .

§22. The equations of hydrodynamics. The viscosity and thermal conductivity of gases

We have seen that Boltzmann's kinetic equation allows one to obtain, as a consequence, the laws of mechanics of continuous media. However, to actually find the stress tensor σ_{ik} a knowledge of the distribution function f is required.

We pass on to a calculation of the non-equilibrium distribution function for an ideal gas performing a macroscopic motion.

We shall suppose that the velocity of the macroscopic motion of the gas, \mathbf{u} , changes from point to point. However, we shall assume this change to be sufficiently slow, where a slow spatial change of the velocity \mathbf{u} is understood to mean the following: Gas volumes of spatial extent of the order of several mean free path lengths can be considered to be moving with a common constant velocity. As was stated in §20, a local Maxwell distribution is estab-

lished in such volumes. The velocities of motion of different gas volumes can be different, so that $\mathbf{u} = \mathbf{u}(\mathbf{r}, t)$.

We shall confine ourselves to isothermal modes of motion of the gas, so that the temperature is constant throughout the volume of the gas. Assuming that local equilibrium is established, we substitute the value of $f^{(0)}$ into the expression for the stress tensor σ_{ik} and the heat flux q_k .

Formula (18.12) gives

$$\begin{aligned}\sigma_{ik} &= mN \int (v_i - u_i)(v_k - u_k) f^{(0)} dv = \\ &= \rho \overline{(v_i - u_i)(v_k - u_k)} = NkT\delta_{ik} = p\delta_{ik}.\end{aligned}\quad (22.1)$$

Thus the stress tensor reduces to the normal pressure. Analogously, from (18.17) we find

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_k} (\rho u_k) = 0, \quad (22.2)$$

In this approximation eqs. (18.9) and (18.14) for a continuous medium assume the form

$$\rho \left(\frac{\partial u_i}{\partial t} + u_k \frac{\partial u_i}{\partial x_k} \right) = - \frac{\partial p}{\partial x_i}. \quad (22.3)$$

Eq. (22.3) represents Euler's equation. As is well-known, (22.3) is the equation of motion of an ideal fluid.

The equation for the entropy per unit volume (19.1) can be written in the form

$$\frac{\partial s}{\partial t} + \frac{\partial j_k^s}{\partial x_k} = 0. \quad (22.4)$$

This last equation shows that when a fluid is moving its specific entropy remains constant, i.e. the process of displacement is of an adiabatic character.

In this approximation a continuous medium can be considered as an ideal fluid with the equation of state

$$p = NkT. \quad (22.5)$$

The set of equations (22.3), (22.4) and (22.5) determines completely the motion of a gas in the continuous medium approximation. In order to obtain the equations of hydrodynamics of a real (viscous) gas in the same approximation, we shall try to find the solution of Boltzmann's equation by the Chapman-Enskog method of successive approximations. Namely, we shall set

$$f = f^{(0)}(1 + \varphi), \quad (22.6)$$

where $\varphi \ll 1$, and $f^{(0)}$ is the local equilibrium distribution.

For ease of calculation we shall consider the gas to be isothermal and incompressible, external forces to be absent and the pressure to be constant. (To this there corresponds, for example, the motion of a gas between two plane surfaces, one moving and the other stationary.)

The mean velocity, u_i , of the gas can depend on coordinates as well as on time. Substituting (22.6) into Boltzmann's equation and confining ourselves to quantities of the first order of magnitude we find

$$\frac{\partial f^{(0)}}{\partial t} + v_k \frac{\partial f^{(0)}}{\partial x_k} = f^{(0)} I(\varphi), \quad (22.7)$$

where

$$I(\varphi) = \int f_1^{(0)} v_{\text{rel}} \sigma [\varphi_2 + \varphi_3 - \varphi_1] dv_1 d\Omega. \quad (22.8)$$

On the left-hand side we have dropped terms proportional to φ , since the derivatives of $f^{(0)}$ are themselves quantities of the first order of magnitude.

Let us calculate the derivatives on the left-hand side of (22.7), making use of expression (20.8) for $f^{(0)}$. Since N and T are constant quantities, we have

$$\frac{\partial f^{(0)}}{\partial t} + v_k \frac{\partial f^{(0)}}{\partial x_k} = \frac{m}{kT} (v_i - u_i) f^{(0)} \frac{\partial u_i}{\partial t} + \frac{m}{kT} (v_i - u_i) f^{(0)} v_k \frac{\partial u_i}{\partial x_k}.$$

For $\partial u_i / \partial t$ use can be made of the ideal-fluid approximation, i.e. by virtue of (22.3) one can write

$$\frac{\partial u_i}{\partial t} = -u_k \frac{\partial u_i}{\partial x_k}.$$

Hence, finally,

$$\frac{\partial f^{(0)}}{\partial t} + v_k \frac{\partial f^{(0)}}{\partial x_k} = \frac{m}{kT} (v_i - u_i) (v_k - u_k) f^{(0)} \frac{\partial u_i}{\partial x_k}. \quad (22.9)$$

For an incompressible fluid $\partial u_k / \partial x_k = 0$. In this case, instead of (22.9) one can write

$$\frac{\partial f^{(0)}}{\partial t} + v_k \frac{\partial f^{(0)}}{\partial x_k} = \frac{m}{kT} (V_i V_k - \frac{1}{3} V^2 \delta_{ik}) \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} \right) f^{(0)}. \quad (22.10)$$

We denote by U_{ik} the deformation rate tensor

$$U_{ik} = \frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i}.$$

Boltzmann's equation then assumes the form

$$\frac{m}{kT} (V_i V_k - \frac{1}{3} V^2 \delta_{ik}) U_{ik} f^{(0)} = f^{(0)} \int f_1^{(0)} v_{\text{rel}} \sigma [\varphi_2 + \varphi_3 - \varphi_1 - \varphi] d\mathbf{v}_1 d\mathbf{\Omega}. \quad (22.11)$$

The function $\varphi(v)$ must satisfy, in addition to eq. (22.11), a subsidiary condition expressing the constancy of the number of particles and of the momentum and energy of the gas as a whole, i.e.

$$\int f d\mathbf{v} = \int f^{(0)} d\mathbf{v} = N,$$

$$mN \int f \mathbf{V} d\mathbf{v} = \mathbf{P} = Nm \int f^{(0)} \mathbf{V} d\mathbf{v} = \text{const},$$

$$E = \frac{1}{2} mN \int f V^2 d\mathbf{v} = \frac{1}{2} mN \int f^{(0)} V^2 d\mathbf{v} = \text{const}.$$

Whence follow the equalities

$$\int f^{(0)} \varphi d\mathbf{v} = 0, \quad (22.12)$$

$$Nm \int f^{(0)} V_i \varphi d\mathbf{v} = 0, \quad (22.13)$$

$$\frac{1}{2} mN \int f^{(0)} V^2 \varphi d\mathbf{v} = 0. \quad (22.14)$$

It is easily seen that the fulfillment of these five equations is the necessary

condition for the existence of solutions of the non-homogeneous integral equation (22.11). Indeed, if the homogeneous integral equation $I(\varphi) = 0$ has the solution $\varphi = \chi(v)$, then the non-homogeneous equation

$$I(\varphi) = A\varphi \quad (22.15)$$

has a solution only if the orthogonality conditions

$$\int \chi(v) \varphi \, dv = 0. \quad (22.16)$$

are fulfilled. Solutions of the homogeneous integral equation are, for example, the functions

$$\chi_1 = e^{-\alpha V^2}; \quad \chi_2 = \mathbf{V} e^{-\alpha V^2}; \quad \chi_3 = V^2 e^{-\alpha V^2}.$$

It is clear that the relations (22.12)–(22.14) express the necessary orthogonality conditions. From the structure of the integral equation (22.11) it is seen that one has to try to find its solution in the form

$$\varphi = \psi_{ik} \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} \right) = \psi_{ik} U_{ik}. \quad (22.17)$$

In substituting (22.17) into (22.11) the quantity U_{ik} on both sides of the equation will cancel. This means that the solution (22.17) is valid for all values of the deformation rate tensor, as is to be expected.

The tensor ψ_{ik} is symmetric. It can also be assumed that $\psi_{ii} = 0$ always, since for $i = k$, $U_{ii} = 0$ and $\varphi = 0$ always. Substituting (22.17) into (22.11), we find

$$\frac{m}{kT} (V_i V_k - \frac{1}{3} V^2 \delta_{ik}) = \int f_1^{(0)} \sigma v_{\text{rel}} [\psi_{ik}^{(2)} + \psi_{ik}^{(3)} - \psi_{ik}^{(1)} - \psi_{ik}] \, dv_1 \, d\Omega. \quad (22.18)$$

The left-hand side of this equation does not depend on the mean velocity and, consequently, the function ψ_{ik} depends only on the components of the relative velocity V_i .

It is obvious that a change of the tensor $(V_i V_k - \frac{1}{3} V^2 \delta_{ik})$ (for example under rotation in velocity space) should not violate eq. (22.18). Hence it follows that for ψ_{ik} one must write

$$\psi_{ik} = \alpha(V) (V_i V_k - \frac{1}{3} V^2 \delta_{ik}), \quad (22.19)$$

satisfying the requirement mentioned. Here $\alpha(V)$ is a certain scalar function of the scalar argument V . The explicit form of this function is obtained from the solution of the integral equation (22.18). To actually obtain the solution, it is necessary to know the dependence of the cross section on velocity and angles. For real molecules, even monatomic ones, the form of this function is unknown. A general idea of the character of the solutions can be obtained by considering the simple, though unrealistic, case where σ is a function only of the scattering angle α but not of the velocity.

In this case (22.18) assumes the form

$$\begin{aligned} \frac{m}{kT} (V_i V_k - \frac{1}{3} V^2 \delta_{ik}) = \\ = \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} N \int e^{-mV_1^2/2kT} \sigma v_{\text{rel}} [\psi_{ik}^{(3)} + \psi_{ik}^{(2)} - \psi_{ik}^{(1)} - \psi_{ik}] d\Omega d\mathbf{v}_1. \quad (22.20) \end{aligned}$$

It is easily seen that the integral term on the right possesses an important feature: if the function in the bracket is a polynomial, then the integral of this function is also a polynomial.

For our simplified model we shall outline the general idea of calculations. If $g = v_{\text{rel}} \sigma(\alpha, v_{\text{rel}})$ is a function of the scattering angle only, we can write

$$g(\alpha) = \left(\frac{2kT}{m} \right)^{\frac{1}{2}} \sigma \frac{v_{\text{rel}} \sigma(\alpha, v_{\text{rel}})}{\sigma \bar{v}} = \left(\frac{2kT}{m} \right)^{\frac{1}{2}} \sigma F(\alpha).$$

Here σ is the total cross section and $F(\alpha) = v_{\text{rel}} \sigma(\alpha, v_{\text{rel}}) / \sigma \bar{v}$.

It is possible to present the solution of (22.20) in the form*

$$\psi_{ik} = \frac{8}{3\sigma' N} \left(\frac{m}{2kT} \right)^{\frac{1}{2}} (V_i V_k - \frac{1}{3} V^2 \delta_{ik}), \quad (22.21)$$

where

$$\sigma' = \sigma \int F(\alpha) \sin^3 \alpha d\alpha.$$

Correspondingly, the distribution function in the first approximation takes

* See G. Uhlenbeck and G. Ford, *Lectures in statistical mechanics* (American Mathematical Society, Providence, 1963).

the form

$$f = f^{(0)} \left[1 - \frac{8}{3} \left(\frac{m}{2kT} \right)^{\frac{3}{2}} \frac{1}{N\sigma'} (V_i V_k - \frac{1}{3} V^2 \delta_{ik}) U_{ik} \right]. \quad (22.22)$$

By means of this distribution function one can find an explicit expression for the stress tensor, which is just our goal. Substituting the distribution function from (22.22) into (18.12) and calculating the integral, we find

$$\sigma_{ik} = -p\delta_{ik} + \frac{2m}{3\sigma'} \left(\frac{2kT}{m} \right)^{\frac{1}{2}} \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} \right). \quad (22.23)$$

According to (8.12), σ_{ik} is expressed in terms of the tensor U_{ik} and viscosity η .

Comparing (8.12) and (22.23), we arrive at the expression for the viscosity of an ideal gas (in our model $\sigma = \sigma(\alpha)$):

$$\eta = \frac{2m}{3\sigma'} \left(\frac{2kT}{m} \right)^{\frac{1}{2}}. \quad (22.24)$$

In the so-called rigid-sphere model, in which it is assumed that the cross section is equal to the geometrical area of the cross section of the sphere, an analogous expression is obtained for the viscosity. Formula (22.24) is in qualitative agreement with experimental data, although, of course, it cannot claim a quantitative meaning. The agreement of formula (22.24) with experiment is apparently associated with the weak dependence of the distribution function φ on the law of intermolecular interaction. We note that the viscosity of the gas turns out to be independent of the density.

In the general case, without specifying the form of the cross section, one can write for η

$$\eta = -mN \int (v_i - u_i)(v_k - u_k) f^{(0)} \psi_{ik} d\mathbf{v},$$

where ψ_{ik} is expressed by formula (22.19).

Thus we have calculated the viscosity, the first of the kinetic coefficients to be calculated.

Making use of expression (22.23) for σ_{ik} , and substituting it into (18.14), one can write

$$\rho \left(\frac{\partial u_i}{\partial t} + u_k \frac{\partial u_i}{\partial x_k} \right) = -\frac{\partial p}{\partial x_i} + \eta \frac{\partial^2 u_i}{\partial x_k^2} + f_i \quad (22.25)$$

or, in vector form,

$$\rho \frac{\partial \mathbf{u}}{\partial t} = -\nabla p + \eta \nabla^2 \cdot \mathbf{u} + \mathbf{f}. \quad (22.25')$$

This represents the Navier–Stokes equation. We recall that it describes the motion of a viscous incompressible fluid and is applicable to the case of relatively rarefied gases as well as to the case of liquids.

The previous consideration shows that the Navier–Stokes equation can be obtained theoretically for the case of sufficiently rarefied gases, and also that the viscosity η can then be calculated, at least to within a numerical coefficient. We shall not dwell on analogous calculations for a compressible gas*.

In a completely analogous way one can calculate the heat flow in a thermally non-uniform gas.

Assuming that the temperature T of the gas changes from point to point, and that the gas performs no macroscopic motion, we shall again try to find the solution of Boltzmann's equation by the Chapman–Enskog method in the form (22.6). Then in the local equilibrium distribution $f^{(0)}$ we set the mean velocity u_i equal to zero, so that

$$f^{(0)} = N(x_k) \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} e^{-mv^2/2kT} = e^{(\mu - \frac{1}{2}mv^2)/kT}. \quad (22.26)$$

Here the chemical potential μ and the temperature T change from point to point, i.e. depend on the coordinates x_k . In order that the Chapman–Enskog method may be used, it is necessary, however, to consider this change to be sufficiently slow. We shall formulate below the quantitative criterion for a change to be considered slow. Introduction of the chemical potential μ into the local equilibrium distribution is associated with the following simple consideration. In a non-isothermal gas the number of particles per cm^3 , as well as the temperature, varies from point to point. If, however, the gas is at rest, then the pressure in it must be constant. Otherwise no mechanical equilibrium would be possible, and a macroscopic motion would arise in the gas.

If the local equilibrium distribution is written in the form (22.6) and the chemical potential μ is assumed to be a function of the pressure and temperature $\mu = \mu(T, p)$, then μ will depend on the coordinates only through the temperature.

* See, for example, L.D. Landau and E.M. Lifshitz, *Fluid mechanics* (Pergamon Press, London, 1959).

The function f must satisfy eq. (22.7). Let us calculate its left-hand side, making use of (22.26). Obviously, we have

$$\begin{aligned}
 v_k \frac{\partial f^{(0)}}{\partial x_k} &= v_k \frac{\partial f^{(0)}}{\partial T} \frac{\partial T}{\partial x_k} = v_k \frac{\partial T}{\partial x_k} f^{(0)} \left[\frac{1}{kT} \left(\frac{\partial \mu}{\partial T} \right) - \frac{\mu}{kT^2} + \frac{mv^2}{2kT^2} \right] = \\
 &= v_k \frac{\partial T}{\partial x_k} f^{(0)} \left[-\frac{s}{kT} - \frac{h - sT}{kT^2} + \frac{mv^2}{2kT^2} \right] = \\
 &= -v_k \frac{\partial T}{\partial x_k} f^{(0)} \frac{h - \frac{1}{2}mv^2}{kT^2} = -v_k \frac{\partial T}{\partial x_k} \frac{1}{T} \left(\frac{5}{2} - \frac{mv^2}{2kT} \right) f^{(0)}, \quad (22.27)
 \end{aligned}$$

where h is the heat content per molecule. Hence eq. (22.7) assumes the form (analogous to (23.11))

$$-v_k f^{(0)} \frac{1}{T} \left(\frac{5}{2} - \frac{mv^2}{2kT} \right) \frac{\partial T}{\partial x_k} = f^{(0)} \int f_1^{(0)} \sigma v_{\text{rel}} [\varphi_2 + \varphi_3 - \varphi_1 - \varphi] d\mathbf{v}_1 d\Omega. \quad (22.28)$$

The function φ , as before, must satisfy the conditions (22.12)–(22.14).

From the form of eq. (22.28) it is clear that one has to try to find its solution in the form

$$\varphi = \xi_k \frac{\partial T}{\partial x_k}, \quad (22.29)$$

where the vector ξ_k depends on the velocity \mathbf{v} .

Substituting (22.29) into (22.27) we see that the solution (22.29) is valid for all values of the temperature gradient $\partial T / \partial x_k$, which drops out of the equation. We then obtain

$$-f^{(0)} \frac{\mathbf{v}}{T} \left(\frac{5}{2} - \frac{mv^2}{2kT} \right) = \int f_1^{(0)} \sigma v_{\text{rel}} [\xi_3 + \xi_2 - \xi_1 - \xi] d\mathbf{v}_1 d\Omega. \quad (22.30)$$

Eq. (22.30) represents the equation for the determination of the vector ξ . The only vectorial quantity involved in (22.29) is the velocity vector \mathbf{v} . This means that the direction of the vector \mathbf{v} is the only specified direction in space. Hence the vector ξ must be oriented in this direction, i.e.

$$\xi = \alpha(v) \mathbf{v}, \quad (22.31)$$

where $\alpha(v)$ is a scalar function of the scalar argument v . The function $\alpha(v)$ must satisfy, in addition to the integral equation, the orthogonality conditions (22.12)–(22.14).

The actual form of the scalar function α depends on the form of the scattering cross section. If, in particular, the calculation is carried out for a model in which σ depends only on angles, then on substituting from (22.30) it is easily seen that the polynomial

$$\alpha(v) = -\frac{2}{N\sigma'T} \left(\frac{mv^2}{2kT} - \frac{5}{2} \right) \left(\frac{m}{2kT} \right)^{\frac{1}{2}}$$

satisfies eq. (22.30).

Hence the solution of (22.28) is of the form

$$f = f^{(0)} \left[1 - \left(\frac{mv^2}{2kT} - \frac{5}{2} \right) \frac{2}{\sigma'NT} \left(\frac{m}{2kT} \right)^{\frac{1}{2}} v_k \frac{\partial T}{\partial x_k} \right].$$

Knowing the distribution function, one can find the heat flux

$$q_k = \frac{1}{2} m N \int v_k v^2 f d\mathbf{v} = \frac{5mk}{2\sigma'} \left(\frac{2kT}{m} \right)^{\frac{1}{2}} \frac{\partial T}{\partial x_k}. \quad (22.32)$$

Hence it follows that the thermal conductivity κ is equal to

$$\kappa = \frac{5mk}{2\sigma'} \left(\frac{2kT}{m} \right)^{\frac{1}{2}}. \quad (22.33)$$

It is interesting to note that the relation

$$\frac{\kappa}{\eta} = \frac{15}{4} k = \frac{5}{2} c_v \quad (22.34)$$

does not depend on the unknown quantity $\sigma'g$.

Formula (22.33), just as the formula for viscosity, is in qualitative agreement with experiment. In particular, it correctly gives the temperature dependence of the thermal conductivity. Formula (22.34) has quantitative significance and is in complete agreement with experiment.

§23. Relaxation time

Let us consider in somewhat more detail how a spatially uniform gas in the absence of a field of force gets into an equilibrium state.

We confine ourselves to a gas in a state sufficiently close to an equilibrium state. Then, in accordance with what was stated in §21, we can set

$$f = f^{(0)}(1 + \varphi),$$

where $f^{(0)}$ is the equilibrium distribution function.

Boltzmann's equation assumes the form (21.32). In the absence of an external field, for a uniform gas (21.32) is of the form

$$\frac{\partial \varphi}{\partial t} = \int f_1^{(0)} \sigma v_{\text{rel}} [\varphi_3 + \varphi_2 - \varphi_1 - \varphi] d\mathbf{v}_1 d\Omega. \quad (23.1)$$

We can try to find the solution of this non-homogeneous integro-differential equation in the form

$$\varphi(\mathbf{v}, t) = \sum A_i \psi_i(\mathbf{v}) \exp(-t/\tau_i(v)). \quad (23.2)$$

A linear integral equation is then obtained for $\psi(\mathbf{v})$. To solve it, it is necessary to define the function $\sigma(v; \alpha)$. Thus we see that the time dependence is expressed by a series (23.2) and is defined by an infinite number of quantities τ_i . The convergence of this series and the existence of the set of discrete quantities τ_i have so far not been sufficiently fully investigated. However, the meaning of the series (23.2) and of the quantities τ_i is quite clear. Equilibrium distributions are established in relaxation times τ_i which are different in different velocity intervals.

This means that the Maxwell velocity distributions in different regions of velocity space will be established in different times. It is natural that different mean characteristics of the gas, for example the mean velocity or mean energy, will have their own relaxation periods.

If, however, in a certain rough approximation we disregard this fact and retain only one term of the series, writing

$$f = f^{(0)}(1 + \psi e^{-t/\tau}), \quad (23.3)$$

then for φ we obtain

$$-\frac{\varphi}{\tau} = \frac{I}{f_0}. \quad (23.4)$$

To an order of magnitude, τ can be estimated as

$$\frac{1}{\tau} \simeq N \sigma \bar{v}_{\text{rel}}. \quad (23.5)$$

The factor N arises from the definition of the mean. Furthermore, with the same degree of accuracy we can write

$$\varphi = (f - f^{(0)})/f^{(0)}.$$

In this approximation, the kinetic equation assumes the form

$$-\frac{\partial f}{\partial t} = \frac{f - f^{(0)}}{\tau}. \quad (23.6)$$

This last equation shows that the rate at which the f 's approach equilibrium is greater, the greater the departure from the equilibrium distribution. For sufficiently small departures from equilibrium, the law of change, $\partial f/\partial t$, in the form (23.6) gives the character of the relaxation process with an accuracy sufficient for a number of practical applications. This is apparently associated with the fact that the function $f^{(M)}$ together with the corresponding weight factor has a rather sharp maximum. Hence the major contribution to the relaxation process is given by particles with a velocity close to the mean velocity.

We shall see in what follows that this holds in particular for a gas of fermions.

As well as the kinetic equation (23.6), use is often made of the relaxation approximation for obtaining the distribution function in the general case of a spatially non-uniform gas in a force field. Namely, the kinetic equation is written in the form

$$\frac{\partial f}{\partial t} + v_k \frac{\partial f}{\partial x_k} + \frac{F_k}{m} \frac{\partial f}{\partial v_k} = -\frac{f - f^{(0)}}{\tau}. \quad (23.7)$$

Here τ is considered a function of coordinates, since the gas density ρ changes from point to point.

From the aforesaid it is clear that in this approximation the kinetic equation has only a qualitative meaning. Nevertheless, for small departures from equilibrium, the kinetic equation can often be brought to the form (23.7) under certain assumptions on the character of the function τ . In particular, this holds for spatially uniform gases in an external force field.

To illustrate the degree of accuracy of the relaxation approximation, we shall calculate, in this approximation, the viscosity of an incompressible gas. In the relaxation approximation we have

$$v_k \frac{\partial f}{\partial x_k} = -\frac{f - f^{(0)}}{\tau}. \quad (23.8)$$

For small departures from an equilibrium state, one can substitute expression (22.10) into the left-hand side of (23.8), so that

$$-\frac{m}{kT} (V_i V_k - \frac{1}{3} V^2 \delta_{ik}) U_{ik} f_0 = \frac{f - f_0}{\tau}.$$

Hence

$$f = f_0 - \frac{m\tau}{kT} (V_i V_k - \frac{1}{3} V^2 \delta_{ik}) U_{ik} f_0. \quad (23.9)$$

Comparing this expression with (22.22), we see that they turn out to be identical if we set

$$\tau = \frac{4}{3\sqrt{2}} \frac{1}{Ng} \simeq \frac{l}{v_{\text{rel}}}, \quad (23.10)$$

where $l = 1/N\sigma_{\text{tot}}$ is the mean free path.

Thus if the scattering cross section is considered to be independent of the velocity, the relaxation approximation gives the same result as the exact solution of the linearized Boltzmann's equation.

If the cross section depends on the velocity, then the difference in the expression for the perturbed distribution function will amount to a numerical coefficient. Correspondingly, the numerical coefficients in the expression for the viscosity will be different. Since the form of the functional dependence on velocity and angles of the cross section for molecular collisions is not well known, there is no certitude as to the values of the numerical coefficients in the formulae for the kinetic coefficients. Therefore the relaxation approximation, which considerably simplifies the calculations, yields in essence an equally accurate solution of the problem.

§24. The diffusion of an admixture of a gas of light particles into a gas of heavy particles

A very important case, allowing substantial simplification of Boltzmann's equation, is the diffusion of gas particles of low mass into a basic gas of heavy particles.

We shall assume that the concentration of the admixture to the basic gas is small, i.e. that the number of gas particles of low mass per unit volume $n \ll N$, where N is the number density of the heavy gas particles. The mass of the light gas particles, m , will be considered small in comparison with the mass M of the particles of the basic gas. Motion of the admixture can occur under the action of a concentration difference (diffusion) or temperature difference (thermal diffusion).

We shall consider the number of particles of the basic gas per unit volume to be constant throughout the gas. The number of particles of admixture per unit volume varies along a certain direction which we shall choose as the x -axis. We shall assume the state of the system to be stationary, and shall not consider the action of an external force field. We shall write the kinetic equation for the distribution function of the light gas particles.

The distribution function varies only in the direction of the x -axis, so that it can be written in the form $f(p, \theta, x)$, where p is the momentum of the particle, and θ is the angle between the momentum vector and the x -axis. Since the number of particles of the admixture is small, their collisions with one another can be disregarded. Hence in the collision integral one need only retain the term taking into account collisions between the particles of the admixture and the particles of the basic gas. Collisions between light and heavy particles can be considered completely elastic, and the velocity of the motion of the light particles can be considered large in comparison with the velocity of motion of the heavy molecules of the basic gas. We shall assume the latter to be at rest and set

$$v_{\text{rel}} \approx v,$$

where v is the velocity of motion of the particles of the admixture. Correspondingly, $\sigma(v_{\text{rel}}, \alpha) \approx \sigma(v, \alpha)$.

The collision integral takes the following form

$$I = \int v_{\text{rel}} \sigma(v_{\text{rel}}, \alpha) (Ff' - Ff) d\mathbf{p}_1 d\Omega \approx vN \int \sigma(v, \alpha) (f' - f) d\Omega. \quad (24.1)$$

Here F denotes the distribution function of the molecules of the basic gas, which is equivalent to the functions f_3 and f_1 in the notation of the preceding section. f' denotes the function f_2 , and $N = \int F d\mathbf{p}_1$ is the total number of molecules of basic gas per unit volume.

Since the particles of the admixture undergo only completely elastic collisions, their momenta \mathbf{p}' and \mathbf{p} respectively before and after a collision have one and the same absolute value; the process of collision is accompanied by a change in the direction of motion, so that

$$f'(\mathbf{p}', x) \equiv f_2(p, \theta', x) = f(\theta', x),$$

$$f(\mathbf{p}, x) \equiv f(p, \theta, x) = f(\theta, x).$$

For brevity we shall drop the argument p . Physically this means that we shall seek the dependence of the distribution of the particles of the admixture on the x -coordinate and on the direction of motion with respect to the x -axis, for a given absolute value of the momentum. As a result, I can be written in the form

$$I = vN \int \sigma(p/m, \alpha) [f(\theta', x) - f(\theta, x)] d\Omega.$$

The kinetic equation can correspondingly be written as

$$\frac{p}{m} \cos \theta \frac{\partial f(\theta, x)}{\partial x} = vN \int \sigma(p/m, \alpha) [f(\theta', x) - f(\theta, x)] d\Omega. \quad (24.2)$$

As distinct from eq. (14.11), the kinetic equation found for the distribution function of the particles of the admixture is a linear integro-differential equation. Its approximate solution can be found in the form of an expansion in powers of a small parameter, in the given case the ratio v_x/v . This method of solving the kinetic equation is called the Lorentz method. If the concentration difference or temperature difference giving rise to a systematic motion of the admixture along the x -axis is sufficiently small (see below), then the ordered motion will be superposed on the random motion. On the average it can be assumed that the velocity of the ordered motion, v_x , is small in comparison with the velocity of the random motion, v . The departure of the system from an equilibrium state, in which there is total isotropy of velocities, will be small. Hence we shall seek the solution of eq. (24.2) in the form

$$f(\theta, x) \approx f_0(x) + v_x f_1(x) + \dots \approx f_0 + v \cos \theta f_1. \quad (24.3)$$

Here $f_0(x)$ is the equilibrium distribution function at point x (the argument p is omitted for brevity), i.e.

$$f_0(p, x) \equiv f_0 = n(x) \frac{m^3}{(2\pi k T_\alpha)^{3/2}} e^{-p^2/2mkT};$$

where $n(x)$ is the number of particles of admixture per unit volume at the point x , and $f_1(x) v \cos \theta$ is the small correction to the statistical distribution which is sought.

The quantity v_x/v or $\cos \theta$, characterizing the degree of anisotropy of the distribution function, is proportional to the value of the concentration difference or temperature difference giving rise to this anisotropy. Hence the expansion (24.3) represents, in essence, an expansion in a series of powers of the small anisotropy of the distribution function.

Substituting expansion (24.3) into (24.2), we note first of all that the equilibrium distribution function reduces the collision integral to zero. Thus the quantity

$$\begin{aligned} I &= v^2 N \int \sigma(v, \alpha) f_1(x) [\cos \theta' - \cos \theta] d\Omega = \\ &= v^2 N f_1(x) \int \sigma(v, \alpha) [\cos \theta' - \cos \theta] \sin \alpha d\alpha d\beta. \end{aligned}$$

will stand on the right-hand side of (24.2). On the left-hand side we have

$$\frac{p_x}{m} \frac{\partial f_0}{\partial x} + \left(\frac{p_x}{m} \right)^2 \frac{\partial f_1}{\partial x} \approx v \cos \theta \frac{\partial f_0}{\partial x}.$$

The term containing the derivative $\partial f/\partial x$ is dropped, since it is small in comparison with the term retained. Thus we obtain

$$\cos \theta \frac{\partial f_0}{\partial x} = v N f_1(x) \int \sigma(v, \alpha) [\cos \theta' - \cos \theta] d\Omega. \quad (24.4)$$

For what follows it is necessary to relate the angles θ' and θ of the momenta of the particles with respect to the x -axis to the scattering angle α . For this we make use of the well-known formula of spherical trigonometry (eq. (1.6) of Part I) writing

$$\cos \theta' = \cos \theta \cos \alpha + \sin \theta \sin \alpha \cos \beta,$$

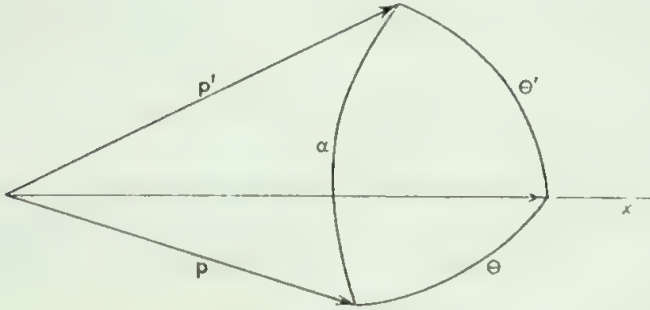


Fig. VI.4.

where $\beta = \psi_1 - \psi_2$, and ψ_1 and ψ_2 are the azimuthal angles of the vectors of momenta \mathbf{p} and \mathbf{p}' (fig. VI.4).

Then we can write

$$\begin{aligned} \int \sigma(v, \alpha) [\cos \theta' - \cos \theta] \sin \alpha \, d\alpha \, d\beta &= \\ &= \int \sigma(v, \alpha) \cos \theta (\cos \alpha - 1) \sin \alpha \, d\alpha \, d\beta + \\ &+ \int \sigma(v, \alpha) \sin \theta \sin \alpha \cos \beta \sin \alpha \, d\alpha \, d\beta = \\ &= -2\pi \cos \theta \int \sigma(v, \alpha) (1 - \cos \alpha) \sin \alpha \, d\alpha. \end{aligned}$$

In the last transformation it is taken into account that $\int \cos \beta \, d\beta = 0$. Then (24.4) assumes the form

$$\cos \theta \frac{\partial f_0}{\partial x} = -\cos \theta \, 2\pi N \left[\int_0^\pi \sigma(v, \alpha) (1 - \cos \alpha) \sin \alpha \, d\alpha \right] f_1 v.$$

Hence it follows that the function f_1 to be determined is equal to

$$f_1 = -\frac{1}{vN\sigma_{tr}} \frac{\partial f_0}{\partial x}, \quad (24.5)$$

where

$$\sigma_{tr} = 2\pi \int_0^\pi \sigma(v, \alpha) (1 - \cos \alpha) \sin \alpha \, d\alpha.$$

The quantity σ_{tr} is called the transport scattering cross section. In the particular case of a cross section independent of the scattering angle, $\sigma(v, \alpha) = \sigma(v)$, we have

$$\sigma_{tr} = \sigma(v) 2\pi \int_0^\pi (1 - \cos \alpha) \sin \alpha d\alpha = 4\pi\sigma(v).$$

If the colliding particles are considered rigid spheres of radii r_1 and r_2 , then $\sigma_{tr} = \pi(r_1 + r_2)^2$.

The quantity

$$\lambda_{tr} = 1/N\sigma_{tr} \quad (24.6)$$

is called the transport mean free path for the motion of light gas particles in a heavy gas. Its obvious meaning will be elucidated below. Introducing into (24.5) the definition of the mean free path, we find

$$f_1(x) = -\lambda_{tr} \frac{\partial f_0(x)}{\partial x} \frac{1}{v} \quad (24.7)$$

and

$$f(p, x, \theta) = f_0(p, x) - \lambda_{tr} \cos \theta \frac{\partial f_0(p, x)}{\partial x}. \quad (24.8)$$

Knowing the distribution function, one can find the mean flux of particles of the admixture in the direction defined (i.e. along the x -axis). By definition, the mean particle flux in the direction of the x -axis, j_x , is equal to

$$j_x = \int v_x f dp.$$

Indeed, $(v_x f)$ gives the number of light particles of a given momentum passing through a cross section of 1 cm^2 per second in the direction of the x -axis. Integrating over all values of the momentum, we find the total number of particles passing through an area of 1 cm^2 per second in the specified direction.

Substituting the value of f from (24.8), we find

$$j_x = \int v \cos \theta f_0(p, x) dp - \int v \lambda_{tr} \cos^2 \theta \frac{\partial f_0}{\partial x} dp. \quad (24.9)$$

In view of the isotropy of the distribution function $f_0(p, x)$ the first integral reduces to zero, so that

$$j_x = - \int v \lambda_{tr} \cos^2 \theta \frac{\partial f_0}{\partial x} p^2 dp \sin \theta d\theta d\psi. \quad (24.10)$$

Let us first consider the case where the gas temperature is constant, and along the x -axis the admixture concentration gradient is $\partial c / \partial x$, where $c = n / (N + n) \approx n / N$. Then the integration can be carried out directly, writing

$$\begin{aligned} j_x &= - \int_0^\pi \int_0^{2\pi} \cos^2 \theta \sin \theta d\theta d\psi \frac{\partial}{\partial x} \int_0^\infty v \lambda_{tr} f_0 p^2 dp = \\ &= - \frac{4}{3} \pi \frac{\partial}{\partial x} \int \frac{v}{N \sigma_{tr}} f_0 p^2 dp = - \frac{1}{3} \left(\frac{\bar{v}}{\sigma_{tr}} \right) \frac{\partial c}{\partial x} = - \frac{1}{3} \left(\frac{\bar{v}}{N \sigma_{tr}} \right) N \frac{\partial c}{\partial x}. \end{aligned} \quad (24.11)$$

Here the bar denotes the mean value over the equilibrium distribution, i.e.

$$\left(\frac{\bar{v}}{\sigma_{tr}} \right) = \frac{1}{n} \int \left(\frac{v}{\sigma_{tr}} \right) f_0 dp = \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \int \frac{v}{\sigma_{tr}} e^{-p^2/2mkt} p^2 dp.$$

The quantity

$$D = \frac{1}{3} \left(\frac{\bar{v}}{N \sigma_{tr}} \right) = \frac{1}{3} (\bar{\lambda}_{tr} \bar{v}) \quad (24.12)$$

is the diffusion coefficient. Finally, for the light gas particle flux in the heavy gas we obtain

$$j_x = -DN \frac{dc}{dx}. \quad (24.13)$$

If instead of the concentration c use is made of the number of particles n of light gas per unit volume, then

$$j_x = -D \frac{\partial n}{\partial x}. \quad (24.14)$$

From the definition of the diffusion coefficient D it is clear that it is an

essentially positive quantity. Hence the flux of particles of the light gas admixture is always oriented in the direction of decreasing light gas concentration. It is just this phenomenon which is called diffusion.

If a constant concentration difference along the x -axis is somehow always maintained, then in the gas mixture there will be a stationary motion of the light component (admixture) in the direction from the higher to the lower concentration. If, on the contrary, the concentration drop is not artificially maintained, then the concentration will become levelled and the composition of the mixture will become constant throughout the container.

We have restricted ourselves to the case of diffusion of an admixture of light gas particles in a gas of heavy particles. Expressions for the diffusion of a heavy gas particle admixture in a gas of light particles can be calculated in an analogous way. We shall not dwell on this here.

Let us now consider in more detail the meaning of the mean free path λ_{tr} .

If a light particle is moving with mean velocity \bar{v} , in one second it collides with all the particles (at rest) confined in a cylinder of height \bar{v} and base area σ_{tr} described by the light particle. The appearance of σ_{tr} instead of σ is equivalent to taking into account the fact that not all the collisions are effective. Thus, for example, collisions with a scattering angle $\alpha \approx 0$ change the direction of flight only a small amount and are less likely to make the particle go out of the cylinder than collisions with $\alpha = \pi$. The total number of collisions is correspondingly equal to $N\bar{v}\sigma_{tr}$. The path traversed on the average between two consecutive collisions, i.e. the mean free path, is equal to

$$\lambda_{tr} = \frac{\bar{v}}{N\bar{v}\sigma_{tr}} = \frac{1}{N\sigma_{tr}}. \quad (24.15)$$

It is useful to note the dependence of λ_{tr} on the pressure in the gas. Since $p = NkT$, we have

$$\lambda_{tr} = kT/p\sigma_{tr}. \quad (24.16)$$

If, in particular, the colliding particles are considered to be rigid spheres and σ_{tr} is expressed by the formula $\sigma_{tr} = \pi(r_1 + r_2)^2$, where r_1 and r_2 are the radii of the particles, then

$$\lambda_{tr} = \frac{kT}{p} \frac{1}{\pi(r_1 + r_2)^2}.$$

Correspondingly, the coefficient of diffusion is

$$D = \frac{1}{3} \lambda_{tr} \bar{v} = \frac{1}{3\pi(r_1 + r_2)^2 p} \left(\frac{8(kT)^3}{\pi m} \right)^{\frac{1}{2}}. \quad (24.17)$$

Thus the coefficient of diffusion depends on the properties of the diffusing particles (their mass and size) as well as on the radius of the molecules of the basic gas. It increases with temperature, $\sim T^{\frac{3}{2}}$, and is inversely proportional to the pressure.

Let us now find the conditions of applicability of the approximate solution found. The solution is applicable if expansion (24.3) converges sufficiently rapidly. For this it is necessary that in its turn the following condition be fulfilled:

$$|v f_1| \ll f_0,$$

or, if the value of f_1 is substituted from (24.5),

$$\lambda_{tr} \frac{1}{f_0} \frac{\partial f_0}{\partial x} \ll 1.$$

This last inequality means that the equilibrium function must change sufficiently little over a distance equal to the mean free path.

§25. Thermal diffusion in gases

Above, in considering the motion of the light gas particles, we assumed the gas temperature to be constant. We shall now give up this assumption and shall consider the more general case where there is a variation in concentration of the diffusing gas together with a temperature variation along the x -axis.

We shall again find the flux of the light gas particles along the x -axis, making use of formula (24.10). However we cannot now take out of the spatial differentiation sign the quantities depending on gas temperature, since the temperature itself varies from point to point. Hence we rewrite formula (24.10) in the form

$$j_x = - \int \frac{\cos^2 \theta}{N} \frac{d}{dx} \left(\frac{v}{\sigma_{tr}} f_0 \right) dp. \quad (25.1)$$

We have brought under the spatial differentiation sign the quantities σ_{tr} and v which are functions of the true velocity of the particles (but not of the mean velocity of the particles, which depends on the temperature).

Carrying out the integration over angles, analogous to formula (24.11), we find

$$j_x = -\frac{4\pi}{3N} \frac{d}{dx} \int \frac{v}{\sigma_{tr}} f_0 p^2 dp = -\frac{1}{3N} \frac{d}{dx} n \left(\frac{\overline{v}}{\sigma_{tr}} \right). \quad (25.2)$$

In order to pass to the usual notation in which j_x is expressed as a function of the light gas particle concentration, we write

$$N = p/kT,$$

where p is the gas pressure, constant throughout the mixture. We then have

$$\begin{aligned} j_x &= -\frac{kT}{3} \frac{d}{dx} \frac{n}{p} \left(\frac{\overline{v}}{\sigma_{tr}} \right) = \\ &= -\frac{kT}{3} \frac{d}{dx} \frac{n}{NkT} \left(\frac{\overline{v}}{\sigma_{tr}} \right) = -\frac{T}{3} \frac{d}{dx} \frac{c}{T} \left(\frac{\overline{v}}{\sigma_{tr}} \right) = \\ &= -\frac{1}{3} \left(\frac{\overline{v}}{\sigma_{tr}} \right) \frac{dc}{dx} - \frac{Tc}{3} \frac{d}{dT} \left[\frac{1}{T} \left(\frac{\overline{v}}{\sigma_{tr}} \right) \right] \frac{dT}{dx}. \end{aligned} \quad (25.3)$$

Introducing the diffusion coefficient, we find

$$j_x = -DN \frac{dc}{dx} - Tc \frac{d}{dT} \left[\frac{N}{T} D \right] \frac{dT}{dx}. \quad (25.4)$$

The formula for the particle flux is usually written in the form

$$j_x = -DN \left(\frac{dc}{dx} + k_T \frac{dT}{dx} \right). \quad (25.5)$$

Comparison of formulae (25.4) and (25.5) gives

$$k_T = c \frac{d}{dT} \ln \left[\frac{1}{T} \left(\frac{\overline{v}}{\sigma_{tr}} \right) \right] \quad (25.6)$$

The quantity k_T is called the coefficient of thermal diffusion (see below).

To elucidate the meaning of the result obtained, we shall set $dc/dx = 0$, i.e. shall consider the case where the concentration of admixture is constant in the container of the gas. Then

$$j_x = -DNk_T \frac{dT}{dx}. \quad (25.7)$$

We see that in the presence of a temperature gradient in a gas mixture of constant composition there arises a motion of the particles of the admixture with respect to the basic gas. This phenomenon is called thermal diffusion or thermodiffusion. The value of the thermodiffusion flux is determined by the magnitude of the temperature gradient and by the value of quantity k_T .

The presence of a thermodiffusion flow, i.e. of relative motion of the particles of an admixture in a definite direction, will cause a change in the composition of the mixture i.e. the appearance of a concentration gradient $dc/dx \neq 0$. This latter effect will in its turn lead to an appearance of a flux of the particles of the admixture, which will reduce the accumulation of the admixture due to thermal diffusion. As a result, a state will be established such that the particle flux due to thermal diffusion is completely compensated for by the flux due to ordinary diffusion. In this case the total flux of admixture particles with respect to the basic gas will be equal to zero, and formula (25.5) gives

$$j_x = -DN \left(\frac{dc}{dx} + k_T \frac{dT}{dx} \right) = 0,$$

so that

$$\frac{dc}{dx} = -k_T \frac{dT}{dx}. \quad (25.8)$$

In a non-isothermal gas mixture, a stationary concentration gradient defined by formula (25.8) is established. In the general case, irrespective of whether or not such a steady state is reached, a temperature gradient gives rise to a concentration gradient in the gas mixture.

We have not, so far, pointed out in which direction the light gas particles will move; the direction of increasing or decreasing temperature. From (25.7) it is clear that the direction of flow is defined by the sign of the coefficient of thermal diffusion k_T , since D is an essentially positive quantity. The sign of k_T depends on that of the derivative $d/dT[(v/\sigma_{tr})/T]$. It is impossible to in-

dicates the sign of this derivative for the general case. It depends on the actual law of interaction between the molecules of the admixture and the basic gas.

The phenomenon of thermal diffusion is of great practical importance. It is used for the separation of gas mixtures, in particular of mixtures of isotopes. In a container with a gas mixture let one of the walls be maintained at temperature T_1 , and another at a temperature $T_2 > T_1$. Then a thermodiffusion flow will arise in the container. As a rule, the light gas particles move against the heat flow, i.e. towards the warmer wall. If a mixture of constant composition was initially in the container, then as a result of thermal diffusion, it will be enriched in one component, for example the light one, at the warmer wall. By extracting the particles of the light component here, one can maintain a stationary thermodiffusion flow and carry out a separation of the light and heavy components of the mixture.

§26. The dispersion of sound

One of the relatively new applications of the kinetic theory of gases is to find the law of sound dispersion in gases. Let us consider an equilibrium monatomic gas in which a plane sound wave is propagating. We choose the direction of propagation to be the x -axis. It is natural to try to find the perturbation of all the quantities characterizing the state of the gas, in particular, the perturbation of the distribution function in the form

$$\varphi = F(v) e^{i(\omega t - \kappa x)}. \quad (26.1)$$

The equation for the perturbed distribution function is of the form of (21.18). Substituting (26.1) into (21.32) gives

$$i\omega F(v) - i v \kappa F(v) = I(\varphi_0). \quad (26.2)$$

We expand $F(v)$ in a series in terms of the eigenfunctions of eq. (21.33)

$$F(v) = \sum \alpha_i \varphi_i^{(0)}.$$

We then have

$$\sum [i\omega \alpha_i \varphi_i^{(0)} - i v \kappa \alpha_i \varphi_i^{(0)}] = \lambda_k \alpha_k \varphi_k^{(0)}. \quad (26.3)$$

Multiplying (26.3) by $\varphi_k^{(0)}$ and integrating, we arrive at the system of linear

algebraic equations

$$i\omega\alpha_k + i\kappa \sum \beta_{ik}\alpha_i - \lambda_k\alpha_k = 0, \quad (26.4)$$

where

$$\beta_{ik} = \frac{\int \varphi_i^{(0)} v \varphi_k^{(0)} dv}{\int \varphi_i^{(0)} \varphi_k^{(0)} dv}.$$

For the system (26.4) to have non-zero solutions, it is necessary that its determinant reduce to zero. This gives

$$\|(i\omega + \lambda_k)\delta_{ik} - i\kappa\beta_{ik}\| = 0. \quad (26.5)$$

Determinant (26.5) is, in general, of infinitely high order. To define the law of dispersion $\omega(k)$, it is necessary to evaluate this determinant and to find its roots.

If we confine ourselves to the first known eigenfunctions of the homogeneous equation (21.32), then we have

$$\varphi_1 = 1, \quad \varphi_2 = v, \quad \varphi_3 = \frac{1}{2}mv^2,$$

$$\lambda_1 = 0, \quad \lambda_2 = 0, \quad \lambda_3 = 0.$$

In this case the first three columns and three rows are to be retained in the determinant.

A simple calculation leads to the formula obtained in the hydrodynamic approximation

$$\omega = \left(\frac{5kT}{3m}\right)^{\frac{1}{2}} \kappa.$$

The velocity of sound c turns out to be constant

$$c = \frac{\omega}{\kappa} = \left(\frac{5kT}{3m}\right)^{\frac{1}{2}}.$$

Taking into account eigenfunctions and eigenvalues of higher order, one can find the law of dispersion as a function of frequency.

It turns out that the velocity of sound in this case increases with frequency, and that at the same time absorption of sound occurs. The absorption coefficient rapidly increases with decreasing wavelength. As the sound wavelength $1/k$ approaches the mean free path, the absorption coefficient increases up to the value $1/k$. This means that the propagation of sound as a periodic perturbation is stopped when $k \sim 1/\lambda$.

§27. The linearized Boltzmann equation for quasi-gaseous systems

The application of Boltzmann's equation is not confined to the case of ideal gases. On the contrary, its most important applications are not associated at all with the consideration of ideal gases. There are a large number of important cases where the kinetic behaviour of a system is similar to that of a gas.

In the most general form the properties of such systems, which we shall call quasi-gaseous systems, can be formulated as follows. Let a system of mutually non-interacting light particles (particles of the first kind) be placed in a medium formed by mutually interacting heavy particles (particles of the second kind). Between particles of the two kinds there is a certain interaction having the character of pair collisions. The system of particles of the first kind can be described by a certain one-particle distribution function $f(\mathbf{r}, \mathbf{v}, t)$, since there is no interaction between these particles. Then, in the general case, one can write for the change of the distribution function

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{F} \cdot \nabla_{\mathbf{p}} f = \left(\frac{df}{dt} \right)_{\text{coll}}, \quad (27.1)$$

where $(df/dt)_{\text{coll}}$ describes the change in the distribution function due to the pair interaction; collisions between particles of the first and second kinds.

Further, we assume that the collisions can be considered completely elastic, and that the state of particles of the second kind does not change in collisions. Then we can write

$$\left(\frac{df}{dt} \right)_{\text{coll}} = IN = N \int \sigma(v, \alpha) v [f_1 - f] d\Omega. \quad (27.2)$$

Here, as distinct from (14.10), v represents the velocity of the light particle, and N is the number of heavy particles per unit volume. As a result, we ob-

tain for the distribution function of the light particles the equation

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{F} \cdot \frac{\partial f}{\partial \mathbf{p}} = N \int \sigma v [f_1 - f] d\Omega, \quad (27.3)$$

which represents a linear integro-differential Boltzmann's equation. A particular case of it is eq. (24.2) describing the diffusion of an admixture of light gas particles.

In the chapters devoted to plasma theory and solid-state theory we shall discuss in detail modern concepts about the behaviour of electrons in plasmas and in solid bodies. Bearing in mind subsequent applications, we shall assume the external forces acting upon particles of the first kind to be Lorentz forces:

$$\mathbf{F} = e \left(\mathbf{E} + \frac{1}{c} [\mathbf{v} \times \mathbf{H}] \right).$$

We shall consider the system to be in a non-equilibrium but stationary state, and the temperature and the chemical potential to vary in space, so that $T = T(\mathbf{r})$, $\mu = \mu(\mathbf{r})$.

Then eq. (27.3) takes the form

$$\frac{\mathbf{p}}{m} \cdot \frac{\partial f}{\partial \mathbf{r}} + e \left(\mathbf{E} + \frac{1}{c} [\mathbf{v} \times \mathbf{H}] \right) \cdot \frac{\partial f}{\partial \mathbf{p}} = N \int \frac{\mathbf{p}}{m} \sigma (f_1 - f) d\Omega. \quad (27.4)$$

To find the solution of this stationary linear Boltzmann equation, we shall assume that the local equilibrium distribution $f^{(0)}(\epsilon, T, \mu)$ is defined, where ϵ is the energy of the particle, and μ and T are local values of the temperature and chemical potential. The local equilibrium distribution $f^{(0)}$ can be a Maxwell distribution, a Fermi distribution, or a Bose distribution. Assuming the departure from an equilibrium state to be sufficiently small, one can try to find the solution of (27.4) in the form

$$f \simeq f^{(0)} + f'(\mathbf{p}, \mathbf{r}),$$

where $|f'| \ll f^{(0)}$. Physically this means that all the fields acting on the particles as well as all the temperature gradients and concentration gradients are small. As distinct from $f^{(0)}$, the part f' of the distribution function does not have spherical symmetry in momentum space.

Conservation of the total number of particles requires that the following

condition be fulfilled:

$$\int f'(\mathbf{p}, \mathbf{r}) d\mathbf{p} = 0. \quad (27.5)$$

In the general case one can write

$$f^{(0)} = f^0((\epsilon - \mu)/kT). \quad (27.6)$$

Substituting (37.5) into (27.4), we shall retain only quantities of higher order. Obviously, we have

$$\frac{\mathbf{p}}{m} \cdot \frac{\partial f}{\partial \mathbf{r}} \simeq \frac{\mathbf{p}}{m} \cdot \frac{\partial f^{(0)}}{\partial \mathbf{r}}.$$

We transform the derivative

$$\frac{\partial f^{(0)}}{\partial \mathbf{r}} = \frac{\partial f^{(0)}((\epsilon - \mu)/kT)}{\partial \mathbf{r}} = \frac{\partial f^{(0)}}{\partial \epsilon} (\epsilon - \mu) \frac{\partial}{\partial \mathbf{r}} \left(\frac{1}{kT} \right) - \frac{\partial f^{(0)}}{\partial \epsilon} \frac{1}{kT} \frac{\partial \mu}{\partial \mathbf{r}},$$

so that

$$\frac{\mathbf{p}}{m} \cdot \frac{\partial f^{(0)}}{\partial \mathbf{r}} = \frac{\mathbf{p}}{m} \cdot \frac{\partial f^{(0)}}{\partial \epsilon} \left[(\epsilon - \mu) \nabla \frac{1}{kT} - \frac{\nabla \mu}{kT} \right]. \quad (27.7)$$

Further, we have

$$e\mathbf{E} \cdot \frac{\partial f}{\partial \mathbf{p}} \simeq e\mathbf{E} \cdot \frac{\partial f^{(0)}}{\partial \epsilon} \frac{\partial \epsilon}{\partial \mathbf{p}} = \frac{e\mathbf{p} \cdot \mathbf{E}}{m} \left(\frac{\partial f^{(0)}}{\partial \epsilon} \right), \quad (27.8)$$

$$\begin{aligned} \frac{e}{mc} [\mathbf{p} \times \mathbf{H}] \cdot \frac{\partial f}{\partial \mathbf{p}} &= \frac{e}{mc} \frac{\partial f^{(0)}}{\partial \epsilon} \frac{\mathbf{p}}{m} \cdot [\mathbf{p} \times \mathbf{H}] + \frac{e}{mc} \frac{\partial f'}{\partial \mathbf{p}} \cdot [\mathbf{p} \times \mathbf{H}] = \\ &= \frac{e[\mathbf{p} \times \mathbf{H}]}{mc} \cdot \frac{\partial f'}{\partial \mathbf{p}}, \end{aligned} \quad (27.9)$$

since the first term identically reduces to zero.

Finally, for the collision integral we find

$$I = N \int \frac{p}{m} \sigma [f'_1 - f'] d\Omega,$$

since the equilibrium distribution reduces I to zero.

Substituting (27.7), (27.8) and (27.9) into (27.4), we find finally

$$\begin{aligned} \frac{\mathbf{p}}{m} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\partial f^{(0)}}{\partial \epsilon} \frac{\mathbf{p}}{m} \cdot \left\{ (\epsilon - \mu) \nabla \frac{1}{kT} - \frac{\nabla \mu}{kT} + e\mathbf{E} \right\} = \\ = - \frac{e}{mc} [\mathbf{p} \times \mathbf{H}] \cdot \frac{\partial f'}{\partial \mathbf{p}} + N \int \sigma \frac{p}{m} [f'_1 - f'] d\Omega. \end{aligned} \quad (27.10)$$

Because of the presence of additional terms on the right-hand side, the solution of eq. (27.9) in a magnetic field has some singularities. We shall consider this in ch.3, and in the meanwhile we set $H = 0$. Then eq. (27.10) goes over into a linear integral equation which is conveniently written in the reduced form

$$\mathbf{A} \cdot \mathbf{p} = N \frac{p}{m} \int \sigma \left(\left| \frac{\mathbf{p}}{m} \right|, \alpha \right) [f'_1 - f'] d\Omega.$$

We shall try to find its solution as in §24 in the form

$$f' = -\alpha p A \cos \theta,$$

where $\theta = (\widehat{\mathbf{p}, \mathbf{A}})$, and α is a certain constant. We then have

$$p A \cos \theta = \alpha \frac{p}{m} N A \int \sigma \left(\left| \frac{\mathbf{p}}{m} \right|, \alpha \right) [\cos \theta_1 - \cos \theta] d\Omega,$$

where $\theta' = (\widehat{\mathbf{p}_1, \mathbf{A}})$. The value of $\cos \theta'$ can be expressed in terms of $\cos \theta$ and $\cos \alpha$ by means of a formula of spherical trigonometry (Vol. 1, (I.7)). Using this expression, we find

$$\int \sigma(|\mathbf{p}/m|, \alpha) (\cos \theta \cos \alpha - \cos \theta) d\Omega = -\cos \theta \sigma_{\text{tr}}.$$

Thus we have

$$\alpha = 1/Nv\sigma_{\text{tr}} = \tau.$$

Finally, we obtain for the distribution function

$$f = f^{(0)} - \tau \frac{\mathbf{p}}{m} \frac{\partial f^{(0)}}{\partial \epsilon} \cdot \left\{ (\epsilon - \mu) \nabla \frac{1}{kT} - \frac{\nabla \mu}{kT} + e\mathbf{E} \right\}. \quad (27.11)$$

We note, first of all, that the collision integral in the same approximation can be written in the form

$$NI = N(f - f^{(0)})/\tau.$$

This last expression shows at once that τ represents none other than the relaxation time (compare with (23.6)).

The violation of spherical symmetry of the distribution function in momentum space turns out to be proportional to the value of $\cos \theta$. As a matter of fact, the representation of the distribution function in the form (27.11) expresses its expansion in a series in terms of Legendre polynomials, in which the first term of the expansion is retained. It is easily seen that the subsidiary condition (27.5) on the distribution function is automatically satisfied.

Violation of spherical symmetry of the distribution function gives rise to a mean particle flux whose density is equal to

$$\mathbf{j} = \int \mathbf{v} f d\mathbf{p} = - \int \tau \frac{\partial f^{(0)}}{\partial \epsilon} \left\{ \frac{\mathbf{p}}{m}, (\epsilon - \mu) \nabla \frac{1}{kT} - \frac{\nabla \mu}{kT} + e\mathbf{E} \right\} \mathbf{v} d\mathbf{p}. \quad (27.12)$$

The flux \mathbf{j} , according to (21.8), represents the first moment of the distribution.

It is useful to note that the flux \mathbf{j} depends on the properties of the system of particles only through the derivative of the equilibrium distribution function (27.6) with respect to energy. It is clear that if the energy distribution were uniform, the particle flux would reduce to zero.

In ch. 5 we shall make use of the expressions obtained for f and \mathbf{j} .

§28. The solution of Boltzmann's equation for a quasi-gaseous system in an external force field

In what follows we shall use the general solution of the linearized Boltzmann equation in an external force field.

Let us consider the equation

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{v}} = I. \quad (28.1)$$

We assume the external force field to be weak and seek the solution of (28.1) in the form of a series of successive approximations. Confining ourselves to the zero and first order terms of the expansion, we can write

$$f = f^{(0)}(\mathbf{v}, t) + f'(\mathbf{v}, t), \quad (28.2)$$

where $f^{(0)}$ corresponds to the absence of a field, and f' is proportional to the field.

In order to avoid cumbersome formulae, we first restrict ourselves to a spatially uniform system, i.e. we set $\partial f / \partial \mathbf{r} = 0$.

Then we have

$$\frac{\partial f'}{\partial t} - I(f') = -\frac{\mathbf{F}}{m} \cdot \frac{\partial f^{(0)}}{\partial \mathbf{v}}, \quad (28.3)$$

where $I(f')$ is the value of the collision operator in which the function f' is taken as the distribution function.

Besides eq. (28.3), we introduce Green's equation

$$\frac{\partial W}{\partial t} - I(W) = \delta(\mathbf{v} - \mathbf{v}') \delta(t - t'), \quad (28.4)$$

where W is the Green's function satisfying the conditions

$$W = \begin{cases} 0 & t < t', \\ \infty & t = t', \end{cases} \quad (28.5)$$

$$\frac{\partial W}{\partial t} - I(W) = 0 \quad t > t'.$$

The meaning of the Green's function is very simple: it represents the probability that a particle which at the instant of time t' had a velocity \mathbf{v}' will at the instant of time t have a velocity \mathbf{v} .

The solution of eq. (28.3) can be expressed in terms of Green's function

by the usual formula

$$\begin{aligned} f(\mathbf{v}, t) &= - \int d\mathbf{v}' \int dt' \frac{\mathbf{F}}{m} \cdot \frac{\partial f^{(0)}}{\partial \mathbf{v}'} W(\mathbf{v}, \mathbf{v}', t - t') = \\ &= - \int d\mathbf{v}' \int_{-\infty}^0 dt' W(\mathbf{v}, \mathbf{v}', t - t') \frac{\mathbf{F}}{m} \cdot \frac{\partial f^{(0)}}{\partial \mathbf{v}'} . \end{aligned} \quad (28.6)$$

Here we have made use of the first property of the Green's function (28.5). Formula (28.6) gives the general relation between the change of the distribution function and the transition probability W . Knowing $f'(\mathbf{v}, t)$, one can find the particle flux by the formula

$$\mathbf{j} = \int \mathbf{v} f(\mathbf{v}) d\mathbf{v} = \int \mathbf{v} d\mathbf{v} \int d\mathbf{v}' \int_{-\infty}^0 dt' \frac{\mathbf{F}}{m} \cdot \frac{\partial f^{(0)}}{\partial \mathbf{v}'} W(\mathbf{v}, \mathbf{v}', t - t') . \quad (28.7)$$

Analogous expressions can be obtained for a spatially non-uniform but stationary system. Then instead of (28.3) and (28.4) we have (for simplicity, in the one-dimensional case):

$$v \frac{\partial f'}{\partial x} - I(f') = - \frac{F}{m} \frac{\partial f^{(0)}}{\partial v} , \quad (28.8)$$

$$\frac{\partial W(x, x', t - t')}{\partial x} - \frac{I(W)}{v} = \delta(x - x') \delta(v - v') , \quad (28.9)$$

where W represents the probability that a particle which was at point x' and had velocity v' will pass to point x and acquire velocity v in unit time.

The solution of (28.8) has the form

$$f'(v, r) = - \int W \left(\frac{F}{m} \frac{\partial f^{(0)}}{\partial v'} \right) dx dv' . \quad (28.10)$$

In particular, if one sets $I = W/\tau$ then (28.9) assumes the form

$$\frac{\partial W}{\partial x} + \frac{W}{\tau v} = \delta(x - x') \delta(v - v') .$$

Hence it follows that

$$W = \begin{cases} e^{(x-x')/\tau v}, & x > x', \\ 0, & x < x', \end{cases}$$

and (28.10) can be written in the form

$$f'(v, x) = \int \left(\frac{F}{m} \frac{\partial f^{(0)}}{\partial v'} \right) e^{-x''/\tau v} dx'' . \quad (28.11)$$

The integration is carried out over a segment of trajectory passing through point x in the direction of the field, for which $x'' < x$ (we have made the substitution $(x - x') \rightarrow x''$).

The meaning of the exponential factor becomes very simple if it is written in the form $e^{-x''/\lambda}$, where λ is the mean free path length.

Thus the linearized Boltzmann equation in an external field allows a solution expressed in the most general form in terms of the transition probability W . In what follows we shall use this form of solution.

§ 29. The kinetic equation for polyatomic gases

So far we have confined ourselves to the case where gas molecules have only translational degrees of freedom. This is a good approximation for the treatment of monatomic gases. However, in the more important case of polyatomic gases the applicability of this approximation is not *a priori* obvious. It turns out that it is possible to formulate, in a most general way, a kinetic equation for diatomic (linear) molecules or molecules of the type of a symmetric top*. Such molecules have rotational degrees of freedom. Rotational levels are always strongly excited (see § 44 of Part III), so that they can be considered classically.

It can be assumed that oscillatory degrees of freedom are not excited at temperatures which are not too high. Thus the motion of the molecule is defined by three translational and two rotational degrees of freedom. The rotational state of the molecule can be characterized by two generalized coordinates (for example by two angles) and by two generalized momenta corresponding to these coordinates.

* In this section we follow the studies of Y. Kagan et al., Soviet Physics JETP 14 (1962) 1096, 604; 24 (1967) 1272.

However, it is more convenient to characterize rotational motion by four quantities; three momentum components M_i ($i = 1, 2, 3$) and the angle ψ characterizing the orientation of the molecule in a plane perpendicular to the vector M_i . In these variables the kinetic equation assumes the form

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{v}} + \dot{\psi} \frac{\partial f}{\partial \psi} + \dot{\mathbf{M}} \cdot \frac{\partial f}{\partial \mathbf{M}} = I. \quad (29.1)$$

The collision integral is of the form

$$\begin{aligned} I &= \int (f_2 f_3 w' - f f_1 w) d\mathbf{v}_1 d\Omega d\mathbf{M}_1 d\mathbf{M} d\psi = \\ &= \int w (f_2 f_3 - f f_1) d\mathbf{v}_1 d\Omega d\mathbf{M}_1 d\mathbf{M} d\psi, \end{aligned} \quad (29.2)$$

where the transition probabilities w' and w represent the probabilities of direct and inverse transitions.

By virtue of the principle of detailed balance $w' = w$, so that w does not change when \mathbf{M} is replaced by $(-\mathbf{M})$.

Here

$$d\mathbf{M} = M dM d\Omega.$$

The solid angle elements $d\Omega_v$ and $d\Omega_M$ are defined by the orientation of the vectors \mathbf{v} and \mathbf{M} .

The quantity $\dot{\psi}$ represents the velocity of rotation of a molecule in a plane perpendicular to the vector \mathbf{M} . In order of magnitude $1/\dot{\psi} \sim 10^{-13}$ sec, i.e. is of the order of the collision time. The time $1/\dot{\psi}$ is very short in comparison with the time lapse between two consecutive collisions, τ . In order of magnitude we have

$$\frac{\partial f}{\partial t} \sim \frac{\Delta f}{\tau} \ll \frac{\Delta f}{1/\dot{\psi}}. \quad (29.3)$$

Therefore it can be assumed that the term $\dot{\psi} \partial f / \partial \psi$ is the largest term of eq. (29.1). In the first approximation, (29.1) can be written as

$$\dot{\psi} \frac{\partial f}{\partial \psi} = 0. \quad (29.4)$$

This means that the distribution function can be considered to be indepen-

dent of the angle ψ , i.e. one can set

$$\frac{\partial f}{\partial \psi} = 0. \quad (29.5)$$

Then (29.1) takes the form

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial f}{\partial \mathbf{v}} + \dot{\mathbf{M}} \cdot \frac{\partial f}{\partial \mathbf{M}} = I. \quad (29.6)$$

The local equilibrium distribution function $f(\mathbf{v}, \mathbf{M})$ is of the form

$$f^{(0)} = \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \frac{1}{4\pi JkT} e^{-mV^2/2kT} e^{-M^2/2JkT}, \quad (29.7)$$

where, as before, $\mathbf{V} = \mathbf{v} - \mathbf{u}$ is the velocity of translational motion of the molecule with respect to the gas as a whole, and J is the moment of inertia corresponding to \mathbf{M} .

Let us consider the thermal conductivity of a gas. Calculations are carried out according to the same scheme as for a monatomic gas. Setting $\mathbf{u} = 0$, so that

$$f^{(0)} = \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} \frac{1}{4\pi JkT} e^{-mv^2/2kT} e^{-M^2/2JkT}, \quad (29.8)$$

we write the distribution function in the presence of a temperature gradient in the form

$$f = f^{(0)} \left(1 + \xi_k \frac{\partial T}{\partial x_k} \right). \quad (29.9)$$

Carrying out calculations analogous to those performed in §22, we have instead of (22.27)

$$v_k \frac{\partial f^{(0)}}{\partial x_k} = -v_k \frac{\partial T}{\partial x_k} \frac{1}{T} \left(\frac{7}{2} - \frac{mv^2}{2kT} - \frac{M^2}{2JkT} \right) f^{(0)}.$$

Correspondingly, instead of (22.30) we obtain

$$\begin{aligned} \frac{\mathbf{v}}{T} \left(\frac{7}{2} - \frac{mv^2}{2kT} - \frac{M^2}{2JkT} \right) f^{(0)} = \\ = \int f_1^{(0)} w [\zeta_3 + \zeta_2 - \zeta_1 - \zeta] d\mathbf{v}_1 d\Omega d\mathbf{M}_1 d\mathbf{M}. \end{aligned} \quad (29.10)$$

As distinct from eq. (22.30) containing only one vector \mathbf{v} specifying a definite direction, for eq. (29.10) one can construct three vectors: \mathbf{v} , $\mathbf{M}[\mathbf{M} \cdot \mathbf{v}]$ and $[\mathbf{M} \times \mathbf{v}]$ specifying three different directions in space (the quantity \mathbf{M} is a pseudo-vector). Hence instead of (22.31) one can write for the vector $\boldsymbol{\zeta}$ in the most general form

$$\boldsymbol{\zeta} = \mathbf{v}\alpha + \mathbf{M}(\mathbf{M} \cdot \mathbf{v})\beta + [\mathbf{M} \times \mathbf{v}]\gamma, \quad (29.11)$$

where α , β and γ are scalar functions.

However, it is easily seen that this last term must be equal to zero: $\gamma = 0$. Indeed, the left-hand side of eq. (29.10) is invariant under the replacement of \mathbf{M} by $-\mathbf{M}$. On the other hand, the vector $\mathbf{M} \times \mathbf{v}$ changes sign under this replacement, and the integral term also contains the quantities $f^{(0)}$ and \mathbf{M} which are invariant under this replacement. Hence if in (29.10) we retained the last term of the expansion, the left-hand and right-hand sides of the equation would transform under the replacement of \mathbf{M} by $-\mathbf{M}$ according to different laws.

The scalar functions α and β depend on all the scalar arguments which can be constructed from the quantities involved in (29.10). Three such scalar quantities can be constructed: v^2 , M^2 , $(\mathbf{M} \cdot \mathbf{v})^2$. Thus, finally,

$$\boldsymbol{\zeta} = \mathbf{v}\alpha(v^2, M^2, (\mathbf{M} \cdot \mathbf{v})^2) + \mathbf{M}(\mathbf{M} \cdot \mathbf{v})\beta(v^2, M^2, (\mathbf{M} \cdot \mathbf{v})^2). \quad (29.12)$$

The scalar functions α and β must satisfy the integral equation (29.10) and subsidiary conditions (22.12)–(22.14). The heat flow is defined by the formula

$$q_k = N \int \left(\frac{1}{2}mv^2 + \frac{M^2}{2J} \right) f^{(0)} \varphi d\mathbf{v} = -\kappa \frac{\partial T}{\partial x_k}. \quad (29.13)$$

To find concrete expressions for the functions α and β and, hence, to calculate the coefficient of thermal conductivity, it is necessary to make assumptions about the function w .

Such calculations were carried out in the above quoted studies of Kagan et al. in the approximation where collisions can be considered to be elastic and molecules to be rigid spherocylinders (cylinders bounded at the top and bottom by hemispheres). They are too cumbersome to be given here. As a result, for the thermal conductivity one obtains the value

$$\kappa = 1.6mk \left(\frac{kT}{\pi m} \right)^{\frac{1}{2}} \left\{ l^2 \left[\left(\frac{a}{l} \right)^2 + \left(\frac{a}{l} \right) + 0.12 \right] \right\}^{-1}, \quad (29.14)$$

where l is the height of the cylinder, and a is the radius of the hemispheres. Other kinetic coefficients are calculated in an analogous way. The above reasoning is of rather methodological interest, since it shows how one may seek the solutions of integral equation (29.10) for polyatomic molecules possessing rotational degrees of freedom.

A more interesting qualitative effect which manifests itself in diatomic molecules is the dependence of the kinetic coefficients on external fields.

The thermal conductivity and viscosity of paramagnetic gases turn out to depend on the external magnetic field (the Senftleben effect). Analogous phenomena are observed in gases of polar molecules placed in a static electric field. The two phenomena are of the same physical nature: in the presence of magnetic or dipole moments the molecules are oriented by the field. In the kinetic equation one has to retain the term corresponding to the dependence of the distribution function on the moment. As a result of this an additional heat flow or momentum flow arises.

Let us consider, for example, the case of the thermal conductivity of a paramagnetic gas in a magnetic field. The kinetic equation is of the form

$$-v_k \frac{f^{(0)}}{T} \left(\frac{7}{2} - \frac{mv^2}{2kT} - \frac{M^2}{2JkT} \right) \frac{\partial T}{\partial x_k} + \dot{\mathbf{M}} f^{(0)} \frac{\partial \varphi}{\partial \mathbf{M}} = I(\varphi). \quad (29.15)$$

Here $\dot{\mathbf{M}}$ can be written in quasi-classical approximation as

$$\dot{\mathbf{M}} = [\boldsymbol{\mu} \times \mathbf{H}] = \frac{\mu_0 g}{h} [\mathbf{M} \times \mathbf{H}] = \gamma [\mathbf{M} \times \mathbf{H}],$$

where $\boldsymbol{\mu}$ is the magnetic moment, μ_0 is the Bohr magneton, and g is the gyromagnetic ratio. Hence we have

$$-v_k \frac{f^{(0)}}{T} \left(\frac{7}{2} - \frac{mv^2}{2kT} - \frac{M^2}{2JkT} \right) \frac{\partial T}{\partial x_k} + \gamma [\mathbf{M} \times \mathbf{H}] \cdot \frac{\partial \varphi}{\partial \mathbf{M}} = I(\varphi). \quad (29.16)$$

One can again try to find the solution of integral equation (29.16) in the form

$$\varphi = \xi_k \frac{\partial T}{\partial x_k}.$$

For the vector ξ_k we obtain

$$-\frac{v_k}{T} \left(\frac{7}{2} - \frac{mv^2}{2kT} - \frac{M^2}{2JkT} \right) f^{(0)} + \gamma f^{(0)} [\mathbf{M} \times \mathbf{H}]_i \frac{\partial \xi_k}{\partial \mathbf{M}_i} = I(\xi_k). \quad (29.17)$$

In the presence of a magnetic field, ξ_k is defined by the vectors \mathbf{v} , \mathbf{M} and \mathbf{H} . Calculations analogous to those given above, but even more cumbersome, lead to an expression for ξ_k . The expression for the thermal conductivity is found from formula (29.13). In complete agreement with experiment it turns out that the change in the coefficient of thermal conductivity is

$$\Delta\kappa = \kappa - (\kappa)_{H=0} = F(H/p), \quad (29.18)$$

where F is a certain function of the argument H/p , and p is the pressure of the gas. The temperature dependence of this quantity is determined by the specific form of the cross section for collisions of the molecules.

Finally, it turns out that the thermal conductivity in a magnetic field is anisotropic. The ratio of thermal conductivities for $\nabla T \parallel \mathbf{H}$ and $\nabla T \perp \mathbf{H}$ in strong magnetic fields is

$$\frac{\Delta\kappa_{\parallel}}{\Delta\kappa_{\perp}} \rightarrow \frac{3}{2} \quad \text{for } H \rightarrow \infty,$$

and ceases to depend on any parameters.

In the case of gases of non-linear molecules possessing oscillatory as well as rotational degrees of freedom, the formulation of the kinetic equation turns out to be difficult. This is associated with the possibility of transitions between different vibrational and rotational states which arise in collisions. The Boltzmann equation can by no means be formulated and solved without making certain assumptions about the character of energy transitions from translational to vibrational degrees of freedom as well as on the character (in particular on the multiplicity of the expression) of the latter.

In the case of non-degenerate states it is possible to write Boltzmann's equation for internal degrees of freedom. However, its solution can actually be found only in limiting cases where the energy transition mentioned above proceeds without difficulty or, on the contrary, is rather hindered.

A hindered energy transition between translational and internal degrees of freedom gives rise to a specific relaxation process. This in its turn essentially affects the kinetic properties of the gas, in particular the values of transport coefficients.

In its very meaning, Boltzmann's equation applies to rarefied systems with pair interaction between the particles. Hence Boltzmann's kinetic equation allows one to consider the behaviour of only a rather limited range of systems.

Nevertheless, Boltzmann's kinetic equation is of great importance for modern physical kinetics. It makes possible the deduction of a number of general theoretical conclusions on the character of irreversible processes, allows the formulation of general transport equations, and introduces the most important characteristics of the behaviour of a system in an irreversible process, the kinetic coefficients.

For those systems to which Boltzmann's kinetic equation is applicable, one can obtain an exact expression for the transport equation, relaxation times and transport coefficients and, under certain model assumptions, their numerical values.

Of no less importance is the fact that in a number of physical systems and, above all, in plasmas and solid bodies it is possible to describe the behaviour of the system in the form of the motion of a system of quasiparticles whose properties are close to those of an ideal gas. For this reason, the quantum generalization of Boltzmann's kinetic equation plays a very important role in solid-state theory.

Some other examples of the application of Boltzmann's equation to the solution of kinetic problems will be given in subsequent sections.

§30. The moderation of fast neutrons

One of the most fully investigated problems of kinetics is that of the motion of neutral particles or radiation in matter.

Let particles originate in a region which we shall call the source, and then move in matter, undergoing scattering or absorption.

In the kinetic theory of gases such a statement of the problem appears to be somewhat artificial. However, it is encountered very often. As an example we point out, first of all, the problem of the spatial distribution of radiation. If certain radiation sources emit a particular frequency spectrum $S(\omega, \mathbf{r})$, then in passing through matter this spectrum will change. Absorption and scattering will change the intensity as well as the angular and frequency distributions of the radiation.

Another case of the interaction of neutral particles with matter, which has been investigated in great detail, is the passage of neutrons through matter. As is well known, neutrons are usually obtained as fast particles of energies of the order of several hundred thousand or several million eV. On passing into

matter and undergoing collisions with nuclei, the neutrons are moderated. In inelastic collisions their energy decreases substantially at each scattering. If, however, the energy of the neutrons is lower than the first excited level of the nucleus, then no inelastic collisions occur. Subsequent deceleration is associated with elastic collisions between the neutrons and nuclei.

Although the two processes considered are of completely different physical nature, their formal treatment turns out to be rather similar. Namely, it turns out that the distribution of neutrons and photons in configuration space and their momentum distributions are described by identical equations. This is associated with the fact that in both processes one can disregard the interaction of the particles with each other, and that the interaction with nuclei (for neutrons) or with atoms (for photons) has the character of a short-range interaction.

In what follows we shall for concreteness speak of the motion of a neutron flux in matter.

Let a system of neutrons be characterized by the distribution function $f(\mathbf{r}, \mathbf{p}, t)$, where

$$dn = f(\mathbf{r}, \mathbf{p}, t) d\mathbf{p} dV$$

is the number of neutrons of given momentum \mathbf{p} in volume element dV at the instant of time t .

For the distribution function one can write the kinetic equation characterizing its variation in configuration space and in momentum space with time. We shall write it in the form of a balance of the number of particles per second coming into and going out of a phase-volume element $d\Gamma$.

That is, we have

$$\frac{df}{dt} d\Gamma = \left\{ \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} + q(\mathbf{r}, t) \right\} d\Gamma, \quad (30.1)$$

where $q(\mathbf{r}, t)$ is the strength of the neutron source, i.e. $q(\mathbf{r}, t)$ is the number of neutrons arising per cm^3 per second at the point \mathbf{r} .

As a result of collisions, neutrons go out of and come in to the given phase-space element. For neutron energies sufficiently large in comparison with thermal velocities, one can disregard the motion of the nuclei and consider them to be at rest.

Each collision of a neutron of given momentum with a nucleus takes it out

of the volume $d\Gamma$. We write the number of these collisions as

$$-\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \frac{f}{\tau}, \quad (30.2)$$

where τ is the total free time before collision. Let us write the latter in the form

$$\frac{1}{\tau} = \frac{v}{l_t} = v \left(\frac{1}{l_c} + \frac{1}{l_{sc}} \right),$$

where l_t , l_c and l_{sc} are respectively the total mean free path, the mean free path before capture and the mean free path before scattering.

Because of collisions, neutrons which had a different (larger in absolute value) momentum and were elastically scattered by nuclei go into the element $d\Gamma$:

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = n_0 \int v' \sigma(\mathbf{p}, \mathbf{p}_1) f(\mathbf{r}, \mathbf{p}_1, t) d\mathbf{p}_1, \quad (30.3)$$

where $\sigma(\mathbf{p}, \mathbf{p}_1)$ is the cross section for scattering from an initial momentum \mathbf{p} to a final momentum \mathbf{p}_1 , and n_0 is the number of scattering centres (nuclei) per unit volume.

Taking into account (30.2) and (30.3), Boltzmann's equation (30.1) can be written in the form

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{f}{\tau} = n_0 \int v \sigma(\mathbf{p}, \mathbf{p}_1) f(\mathbf{r}, \mathbf{p}_1, t) d\mathbf{p}_1 + q. \quad (30.4)$$

Eq. (30.4) is linear, since the collisions of neutrons with each other are disregarded.

In neutron physics there are two basic problems: (1) to find the neutron energy distribution (if the energy of the neutrons coming from the sources is known), and (2) to find the spatial distribution of the energy.

We shall begin by considering the first problem. Since we shall not be interested in the spatial distribution of neutrons, we integrate eq. (30.4) over all space and write

$$\int f(\mathbf{r}, \mathbf{p}, t) dV = N(\mathbf{p}, t), \quad \int q(\mathbf{r}, t) dV = Q. \quad (30.5)$$

Clearly, we then obtain

$$\frac{\partial N}{\partial t} + \frac{N}{\tau} = n_0 \int v' \sigma N(\mathbf{p}_1, t) d\mathbf{p}_1 + Q. \quad (30.6)$$

The integral $\int (\partial f / \partial \mathbf{r}) dV$ over all space reduces to zero. In order not to complicate the calculation we shall confine ourselves to the case of a source which emits neutrons of energy E_0 steadily. In this case Q does not depend on the time, $Q = Q_0 \delta(E_0) / 4\pi$. Then (30.6) becomes the stationary equation

$$\frac{N}{\tau} = n_0 \int v' \sigma(\mathbf{p}, \mathbf{p}_1) N(\mathbf{p}_1) d\mathbf{p}_1 + \frac{Q_0 \delta(E_0)}{4\pi}. \quad (30.7)$$

It is more convenient to write (30.7) in the form

$$\psi(\mathbf{p}) = \int \psi(\mathbf{p}_1) w(\mathbf{p}, \mathbf{p}_1) \frac{l_t}{l_{sc}} d\mathbf{p}_1 + \frac{Q_0 \delta(E_0)}{4\pi}, \quad (30.8)$$

where the auxiliary function is

$$\psi(\mathbf{p}) = N(\mathbf{p}) / \tau, \quad (30.9)$$

and $w(\mathbf{p}, \mathbf{p}_1)$ is the transition probability normalized to unity.

$$\int w(\mathbf{p}, \mathbf{p}_1) d\mathbf{p}_1 = 1. \quad (30.10)$$

It is obvious that

$$w(\mathbf{p}, \mathbf{p}_1) / l_{sc} = n_0 \sigma(\mathbf{p}, \mathbf{p}_1).$$

It is now necessary to write the explicit value for the transition probability $w(\mathbf{p}, \mathbf{p}_1)$. Since neutrons interact with nuclei only at very small distances, their elastic scattering obeys the laws for elastic collisions and is isotropic. Hence

$$w(\mathbf{p}, \mathbf{p}_1) = \varphi(p) \delta \left(\frac{p_1^2}{2m} - \frac{p^2}{2m} - \frac{(\mathbf{p}_1 - \mathbf{p})^2}{2M} \right), \quad (30.11)$$

where the argument of the δ -function expresses the energy conservation law in a collision with a nucleus at rest, and $\varphi(p)$ is a certain function of the neu-

tron energy $p^2/2m$. M denotes the mass of the nucleus. The value of $\varphi(p)$ can be found from condition (30.10).

Carrying out the integration, we obtain

$$\begin{aligned} & \int \delta \left(\frac{p_1^2 - p^2}{2m} - \frac{(\mathbf{p}_1 - \mathbf{p})^2}{2M} \right) p_1^2 dp_1 \sin \theta d\theta d\varphi = \\ & = \int \delta \left(\frac{1}{2} p_1^2 \left(\frac{1}{m} - \frac{1}{M} \right) - \frac{1}{2} p^2 \left(\frac{1}{m} + \frac{1}{M} \right) + \frac{pp_1 \cos \theta}{M} \right) p_1^2 dp_1 \sin \theta d\theta d\varphi = \\ & = \frac{2\pi M}{p} \int_{p_{\min}}^{p_{\max}} p_1 dp_1 = \frac{\pi M}{p} (p_{\max}^2 - p_{\min}^2). \end{aligned}$$

Here p_{\min} and p_{\max} are the minimum and maximum momenta before collision, which after collision turn into the momentum p .

According to the elastic collision laws (see (43.33) of Part I)

$$p_{\min} = \frac{M - m}{M + m}, \quad p_{\max} = p.$$

Therefore, finally,

$$\int \varphi(\mathbf{p}_1) \delta d\mathbf{p}_1 = \varphi(p) \frac{4\pi m M}{(m + M)^2} p = 1,$$

Hence $\varphi = (M + m)^2 / 4\pi m M^2 p$ and the normalized scattering probability assumes the form

$$w(\mathbf{p}, \mathbf{p}_1) = \frac{1}{4\pi} \left(\frac{M + m}{M} \right)^2 \frac{1}{mp} \delta \left(\frac{p_1^2 - p^2}{2m} - \frac{(\mathbf{p}_1 - \mathbf{p})^2}{2M} \right). \quad (30.12)$$

The problem of integrating eq. (30.8) for the value of $w(\mathbf{p}, \mathbf{p}_1)$ given by formula (30.12) is very complex. We shall therefore confine ourselves to two limiting cases.

Let us first consider the scattering of neutrons on hydrogen nuclei, i.e. the case $m = M$. Then

$$w(\mathbf{p}, \mathbf{p}_1) = \frac{1}{\pi M p} \delta \left(\frac{p_1^2 - p^2 - (\mathbf{p}_1 - \mathbf{p})^2}{2M} \right). \quad (30.13)$$

Substituting (30.13) into (30.8) we note, first of all, that in view of the isotropy of the scattering, the function w does not depend on angles and depends only on the absolute value of the momentum, so that (30.8) assumes the form

$$w(p) = 2 \int_0^p \psi(p_1) \frac{l_t(p_1)}{l_{sc}(p_1)} \frac{p_1 dp_1}{p^2} + \frac{Q_0 \delta(E_0)}{4\pi} \quad (30.14)$$

or, passing to a new variable, the energy, one can write (30.14) in the form

$$\psi(E) = \int_0^E \psi(E_1) \frac{l_t(E_1)}{l_{sc}(E_1)} \frac{dE_1}{E} + \frac{Q_0 \delta(E_0)}{4\pi}. \quad (30.15)$$

This equation can be solved in an elementary way. Differentiating gives

$$\frac{d\psi}{dE} = \left[\frac{l_t(E)}{l_{sc}(E)} - 1 \right] \frac{\psi(E)}{E} + \frac{Q_0}{4\pi} \left[\frac{\delta(E_0)}{E} + \delta'(E_0) \right].$$

The solution of this linear equation is

$$\psi = \frac{Q_0}{4\pi} \frac{l_t(E_0)}{l_{sc}(E_0)} \exp \left\{ - \int_{E_0}^E \frac{l_t(E_1)}{l_{sc}(E_1)} \frac{dE_1}{E_1} \right\}. \quad (30.16)$$

For the energy distribution function, it follows from (30.9) that

$$N(E) = \frac{Q_0}{4\pi} \frac{l_t(E_0)}{l_{sc}(E_0)} \frac{l_t(E)}{v} \exp \left\{ - \int_{E_0}^E \frac{l_t(E_1)}{l_{sc}(E_1)} \frac{dE_1}{E_1} \right\}. \quad (30.17)$$

If there is no absorption $l = l_t = l_{sc}$ and (30.17) is simplified

$$N(E) = \frac{Q_0}{4\pi} \frac{l}{v} \frac{1}{E}. \quad (30.18)$$

Eq. (30.8) is also substantially simplified in the other limiting case, neutron scattering on heavy nuclei. In this case the energy transfer in each collision is small.

To simplify the calculations we shall confine ourselves to the case where the absorption of neutrons can be disregarded in the process of their moderation in a medium. Then the general case of non-monochromatic sources can be considered. In the absence of absorption $l_t = l_{sc}$ and eq. (30.8) assumes the form

$$\psi(\mathbf{p}) = \int \psi(\mathbf{p}_1) w(\mathbf{p}, \mathbf{p}_1) d\mathbf{p}_1 + Q(E)/4\pi. \quad (30.19)$$

Since we are interested only in the energy spectrum of the moderated neutrons, (30.19) can be integrated over all angles. Then it turns out to be convenient to pass from the absolute value of momentum to a new independent variable*

$$u = \ln(E_0/E) = 2 \ln(p_0/p). \quad (30.20)$$

The grounds for the choice of such a variable will be seen in further calculations.

We define the new unknown function $\psi(u)$ by the relation

$$\psi(u) = \int \psi(\mathbf{p}) d\Omega. \quad (30.21)$$

We set $\mu = (\widehat{\mathbf{p}}, \widehat{\mathbf{p}}_1)$ and

$$w d\mathbf{p}_1 = \eta(u, \mu) du d\Omega.$$

A simple transformation gives

$$\begin{aligned} \eta(u, \mu) &= \eta(u, \mathbf{p}_1, \mathbf{p}) = \\ &= \frac{(M+m)^2}{8\pi Mn} e^{-u} \delta \left(\cos \mu - \frac{M+m}{2m} e^{-\frac{1}{2}u} + \frac{M-m}{2m} e^{\frac{1}{2}u} \right). \end{aligned}$$

Integrating (30.19) over all angles, we find

$$\psi_0(u) = \int_0^u du' \psi_0(u') \int \eta(u - u', \mu) d\Omega. \quad (30.22)$$

* A. Akhiezer and I. Pomeranchuk, *Nekotorye voprosy teorii yadra* (Some problems of nuclear theory) (Gostekhizdat, Moscow, 1950).

The internal integral with the δ -function is easily calculated if it is taken into account that the neutron energy after the elastic collision, E' , lies in the interval (see §43 of Part I)

$$E_{\max} = E \geq E' \geq \left(\frac{M-m}{M+m} \right)^2 E = E_{\min}.$$

Hence introducing the quantity

$$u_M = \ln(E/E_{\min}) = 2 \ln[(M+m)/(M-m)]$$

and integrating over angles in the δ -function, we find

$$\eta_0(u-u') = \int \eta(u-u', \mu) d\Omega = \begin{cases} \frac{M+m}{4mM} e^{-(u-u')} & \text{for } u < u_M, \\ 0 & \text{for } u > u_M. \end{cases} \quad (30.23)$$

Substituting (30.23) into (30.22), we have

$$\psi_0(u) = \int_0^u du' \psi_0(u') \eta_0(u-u') + Q. \quad (30.24)$$

Eq. (30.23) can be solved in closed form by means of a Laplace transformation. Namely, we multiply (30.24) by e^{-zu} and integrate over all u 's in the range $0 \leq u < \infty$. We then find

$$\int_0^\infty \psi_0(u) e^{-zu} du = \int_0^\infty e^{-zu} du \int_0^u du' \psi_0(u') \eta_0(u-u') + \int_0^\infty e^{-zu} Q(u) du. \quad (30.25)$$

Changing the order of integration in the double integral, we obtain

$$\begin{aligned} \int_0^\infty du' \psi_0(u') e^{-zu'} \int_u^\infty e^{-z(u-u')} \eta(u-u') du' &= \\ = \int du' \psi_0(u') e^{-zu'} \int_0^\infty e^{-zu} \eta(u) du & \quad (30.25') \end{aligned}$$

Denoting the Laplace transforms by

$$\tilde{\psi}(z) = \int_0^{\infty} \psi_0(u) e^{-zu} du, \quad (30.26)$$

$$\tilde{\eta}(z) = \int_0^{\infty} \eta(u) e^{-zu} du, \quad (30.27)$$

$$\tilde{Q}(z) = \int_0^{\infty} Q(u) e^{-zu} du \quad (30.28)$$

and taking into account (30.25'), we finally obtain in place of (30.25)

$$\tilde{\Psi}(z) = \tilde{\psi}(z) \tilde{\eta}(z) + \tilde{Q}(z). \quad (30.29)$$

Whence for the transform ψ we find

$$\tilde{\psi}(z) = \frac{\tilde{Q}(z)}{1 - \tilde{\eta}(z)}. \quad (30.30)$$

Inverting the Laplace transformation, we find that the neutron distribution is determined in the form

$$\psi_0(u) = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \tilde{\eta}(z) e^{zu} du = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \frac{\tilde{Q}(z) e^{zu} du}{1 - \tilde{\eta}(z)}. \quad (30.31)$$

Let us find the transform $\tilde{\eta}(z)$. According to (30.25) we have

$$\begin{aligned} \tilde{\eta}(z) &= \frac{(M+m)^2}{4mM} \int_0^{u_M} e^{-u} e^{-zu} du = \frac{(M+m)^2}{4mM} \frac{1 - \exp[-u_M(1+z)]}{1+z} = \\ &= \frac{1}{1 - \exp(-u_M)} \frac{1 - \exp[-u_M(1+z)]}{1+z}. \end{aligned} \quad (30.32)$$

Substituting (30.32) into (30.31), we arrive at a somewhat cumbersome for-

mula for the neutron energy distribution (more precisely, the distribution over the logarithm of the energy):

$$\psi_0(u) = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \frac{\tilde{Q}(z) e^{zu} du}{1 - [1 - \exp(-u_M)]^{-1} [1 - \exp(-u_M(1+z))]} / (1+z) \quad (30.33)$$

The integration is carried out with respect to a straight line parallel to the imaginary axis and lying more to the right than all the poles of the integrand. The contour can be closed by a semicircle of infinitely large radius lying on the left of the straight line. Then the theorem of residues can be applied to the integral. As a rule, $\tilde{Q}(z)$ can be considered a function having no poles. Then the poles of the integrand are points defined by the condition

$$(1 - \exp(-u_M))(1+z) = 1 - \exp(-u_M(1+z)). \quad (30.34)$$

It is obvious that this transcendental equation has a root $z_1 = 0$ as well as an infinite set of roots z_k having a negative real part.

Finding the residue of the integrand of (30.33) and applying the theorem of residues, we get

$$\psi_0(u) = \sum \frac{[1 - \exp(-u_M(1+z_k))] \exp(uz_k) \tilde{Q}(z_k)}{1 - \exp(-u_M) + u_M \exp(-u_M(1+z_k))}. \quad (30.35)$$

For large values of u , i.e. for neutron energies considerably smaller than the energies with which they are emitted by the sources, formula (30.35) is substantially simplified. Since the real value $\operatorname{Re} z_k < 0$, all the terms of the sum (30.35) for $u \gg 1$ are exponentially small in comparison with the first term corresponding to $z_1 = 0$.

Hence instead of (30.35) one can write

$$\psi_0(u) \simeq \frac{1 - \exp(-u_M) \tilde{Q}(0)}{1 - \exp(-u_M) + u_M \exp(-u_M)}. \quad (30.36)$$

By definition (30.28)

$$\tilde{Q}(0) = \int_0^\infty Q(u) du = Q_0,$$

where Q_0 is the total number of neutrons emitted by the source per second. We see from (30.36) that $\psi_0(u)$ does not depend on its argument. It is this simplicity of the distribution function in the variable u that makes this variable convenient.

The neutron energy distribution is of the form

$$N(E) = \frac{1}{v} \psi_0 = \frac{1}{v} \frac{Q_0}{\xi},$$

where the quantity ξ is by definition equal to

$$\xi = \frac{1 - \exp(-u_M) + \exp(-u_M)u_M}{1 - u_M}.$$

This quantity, as shown by simple considerations, represents the mean logarithmic neutron energy loss in one collision.

§31. The spatial distribution of neutrons

Let us now consider the important problem of the neutron spatial distribution.

We assume that the change in neutron energy in scattering on nuclei can be disregarded. This can be done, for example, for neutrons moderated down to an energy $\sim kT$ (i.e. for thermal neutrons). If the energy change is disregarded, then the energy in eq. (30.4) can be considered to be fixed. If the sources emit neutrons into the medium steadily, then a stationary neutron distribution will be established in space. The distribution function can be assumed to depend only on the coordinates and direction of motion of the neutrons. The latter can be characterized by the unit vector

$$\mathbf{v}_1 = \mathbf{v}/v. \quad (31.1)$$

The neutron distribution function satisfies the equation

$$\mathbf{v} \cdot \frac{\partial f(\mathbf{r}, \mathbf{v}_1)}{\partial \mathbf{r}} + \frac{v}{l_t} f(\mathbf{r}, \mathbf{v}_1) = \int \frac{v}{l_{sc}} w(\mathbf{v}_1, \mathbf{v}'_1) f(\mathbf{v}'_1, \mathbf{r}) d\Omega' + \frac{q(\mathbf{r})}{4\pi}. \quad (31.2)$$

The integration in the collision integral is carried out only with respect to the angle, since the absolute value of the momentum does not change in a collision.

Dividing (31.2) by v/l_t and introducing the notation

$$\frac{\alpha}{4\pi} = \frac{l_t}{l_{sc}}, \quad s(r) = q \frac{l_t}{v}, \quad (31.3)$$

we have

$$v_1 l_t \frac{\partial f}{\partial r} + f = \frac{\alpha}{4\pi} \int w(v_1, v'_1) f d\Omega' + \frac{s(r)}{4\pi}. \quad (31.4)$$

Eq. (31.3) is similar to the equation for the distribution function of the molecules of a light gas diffusing in a heavy gas (§24). However, an important difference is the presence of particle sources and the absorption of the particles in the medium.

In the cases where the scattering can be considered to be isotropic $w(v, v') = 1$, eq. (31.4) allows an exact solution. Proceeding from the exact solution of eq. (31.4), one can find the neutron density distribution in space. Comparing it with the distribution obtained in the diffusion approximation found in §24, we can estimate the accuracy of the latter.

To obtain the exact solution, it is convenient to use the method of expansion in a Fourier integral. Writing

$$\psi(k, v_1) = \frac{1}{(2\pi)^3} \int e^{-ik \cdot r} f(r, v_1) dV, \quad \varphi(k) = \frac{1}{(2\pi)^3} \int e^{-ik \cdot r} s(r) dV$$

we have

$$(1 + ik \cdot v_1 l_t) \psi(k, v_1) = \frac{\varphi(k)}{4\pi} + \frac{\alpha}{4\pi} \int \psi(k, v_1) d\Omega'$$

or

$$\psi(k, v_1) = \frac{\varphi(k) + \alpha \int \psi(k, v_1) d\Omega'}{4\pi(1 + ik \cdot v_1 l_t)}. \quad (31.5)$$

Integrating (31.5) over angles, we obtain

$$\int \psi(k, v_1) d\Omega = \frac{\varphi(k) + \alpha \int \psi(k, v_1) d\Omega'}{4\pi} \int \frac{d\Omega}{1 + ik \cdot v_1 l_t}. \quad (31.6)$$

The last integral is directly calculable

$$\int \frac{d\Omega}{1 + i\mathbf{k} \cdot \mathbf{v}_1 l_t} = 2\pi \int_0^\pi \frac{\sin \theta d\theta}{1 + ikl_t \cos \theta} = \frac{2\pi}{ikl_t} \ln \frac{1 + ikl_t}{1 - ikl_t} = \frac{4\pi}{kl_t} \arctan kl_t.$$

Substituting this value into (31.6), we get

$$\int \psi(\mathbf{k}, \mathbf{v}_1) d\Omega = \frac{\varphi(\mathbf{k})}{(kl_t / \arctan kl_t) - \alpha} \quad (31.7)$$

Now we note that the integral on the left-hand side of (31.6) represents none other than the Fourier component of the neutron density

$$\begin{aligned} N(\mathbf{r}) &= \int f(\mathbf{r}, \mathbf{v}_1) d\Omega = \frac{1}{(2\pi)^3} \int \psi(\mathbf{k}, \mathbf{v}_1) e^{i\mathbf{k} \cdot \mathbf{r}} d\Omega d\mathbf{k} = \\ &= \frac{1}{(2\pi)^3} \int e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k} \int \psi(\mathbf{k}, \mathbf{v}_1) d\Omega, \end{aligned} \quad (31.8)$$

where $N(\mathbf{r})$ is the number of neutrons per unit volume of the medium. Substituting (31.7) into (31.8), we find

$$N(\mathbf{r}) = \int e^{i\mathbf{k} \cdot \mathbf{r}} \left(\int \psi(\mathbf{k}, \mathbf{v}_1) d\Omega \right) d\mathbf{k} = \int \frac{\varphi(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k}}{kl_t / \arctan kl_t - \alpha}. \quad (31.9)$$

In order to obtain a concrete expression for the neutron density, let us consider the case of a point source

$$s(\mathbf{r}) = s_0 \delta(\mathbf{r}).$$

Then

$$\varphi(\mathbf{k}) = \frac{l_t}{v} \frac{s_0}{(2\pi)^3}$$

and

$$\begin{aligned}
 N(\mathbf{r}) &= \frac{l_t}{v} \frac{s_0}{(2\pi)^3} \int \frac{e^{ikr \cos \theta} k^2 dk \sin \theta d\theta d\varphi}{kl_t / \arctan kl_t - \alpha} = \\
 &= \frac{l_t}{irv} \frac{s_0}{(2\pi)^2} \int_{-\infty}^{\infty} \frac{e^{ikr} k dk}{kl_t / \arctan kl_t - \alpha}.
 \end{aligned} \quad (31.10)$$

The calculation of the integral can be carried out conveniently in the complex plane. Let us consider the integral over the contour shown in fig. VI.5.

$$I = \int \frac{e^{i\beta z} z dz}{z / \arctan z - \alpha}. \quad (31.11)$$

The integrand in (31.11) has a pole of the first order at the point

$$z_1 / \arctan z_1 = \alpha \quad (31.12)$$

and a branch point

$$z_2 = i.$$

The contour passes by the branch point, the integrals over the large and small circles reduce to zero and there remains only the residue at the point z_1 and the integral with respect to the imaginary axis

$$I = 2\pi i \operatorname{Res}(z_1) + \int_i^{i\infty} \frac{e^{i\beta z} z dz}{z / \arctan z - \alpha}.$$

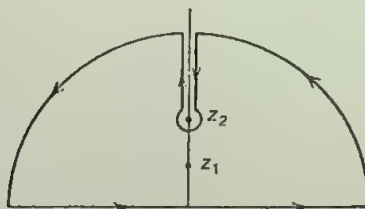


Fig. VI.5

Estimates* show that at large distances from the source ($r \gg l_1$) the second term is small in comparison with the first, and we shall simply drop the integral over the imaginary axis.

We denote the root of eq. (31.12) by i/L , i.e. setting

$$il_1/L = \alpha \arctan il_1/L. \quad (31.13)$$

Calculating the residue at the point (31.13), and substituting the value of I into (31.10), we obtain

$$N(r) = \frac{l_1}{v} \frac{s_0}{2\pi\alpha L^2} \frac{(1 - l_1^2/L^2)}{\alpha + (l_1^2/L^2) - 1} \frac{e^{-r/L}}{\tau}. \quad (31.14)$$

Let us consider the case of weak absorption $l_1 \simeq l_{sc}$. The arc tangent in eq. (31.13) can be expanded in a series, and its approximate solution can be written in the form

$$L^2 \simeq L_0^2 = \frac{1}{3} l_c l_{sc}. \quad (31.15)$$

Obviously, here $l \gg l_{sc}$. Then (31.15) assumes the form

$$N(r) \simeq \frac{3s_0}{4\pi l_{sc} v} \frac{e^{-r/L_0}}{r} \quad \text{for } r \gg l_{sc}. \quad (31.16)$$

It is interesting to compare the spatial particle distribution (31.16) with the particle distribution in the diffusion approximation. In the diffusion approximation the solution of Boltzmann's equation leads, according to (24.17), to the coefficient of diffusion $D = \frac{1}{3} l_{sc} v$. The distribution function $N(r)$ in this case satisfies the equation

$$D \Delta N(r) - \frac{N(r)}{\tau} = \frac{s_0}{4\pi} \delta(r),$$

where $\tau = l_c/v$. Its solution is exactly the same as (31.16) for $l_c \ll l_{sc}$. Thus in the case of weak absorption it turns out that the exact solution of the kinetic equation is, to a high degree of accuracy, the same as the solution of the equa-

* A.D.Galanin, *Teoriya yadernykh reaktorov (Nuclear reactor theory)* (Atomizdat, Moscow, 1957).

tion of the diffusion approximation at large distances from the source $r \gg l_{sc}$. A numerical comparison of the solutions of the kinetic and diffusion equations shows that the two solutions are practically the same at distances from the source larger than $2l_{sc}$.

In the opposite limiting case of very strong absorption, when $l_t \sim l_c$, the equation leads to the value

$$L \simeq 1.05l_t.$$

This means that $N(r)$ decreases substantially at a distance $\sim l_t$ from the source.

In conclusion we stress that the results obtained are of a general character. They equally apply to all particles moving in matter and undergoing scattering and capture, provided the interaction with the scattering centres has the character of a short-range interaction.

§32. The kinetic equation for a plasma disregarding collisions

The direct application of Boltzmann's equation to a plasma calls for a certain amount of caution. Charged particles in a plasma interact according to Coulomb's law, so that the forces of interaction between them decrease relatively slowly with increasing distance. Hence it is necessary to consider those changes which must be introduced into the calculation of the collision integral in order to take into account the specific properties of the Coulomb interaction. However, we shall defer this consideration to §33, and in the meantime we note that although the presence of long-range forces may to a certain degree complicate the problem, it also in one way simplifies it. Namely, since the forces of interaction slowly decrease with increasing distance, collective motions in which relatively large groups of particles take part must arise in a system of charged particles.

One may speak of collective excitations in a plasma in which the system as a whole is involved. In considering such large-scale motions one can disregard non-uniformities (fluctuations) in the system and pair collisions between particles.

From the theory of scattering in a Coulomb force field (see §43 of Part I and §86 of Part V) we know that a considerable deflection of particles occurs for minimum values of the impact parameter

$$\rho_{\min} \sim e^2/mv^2 \sim e^2/\bar{\epsilon},$$

where $\bar{\epsilon} \sim kT$ is the mean energy of particles with temperature T . If the following inequality holds

$$\rho_{\min} \ll \bar{r} \sim n^{-\frac{1}{3}}, \quad (32.1)$$

where \bar{r} is the mean distance between particles and n is their density, then the role of pair collisions becomes relatively unimportant in comparison with the Coulomb interaction of particles at large distances of the order of \bar{r} . In this case the character of the distribution function for a system of particles with Coulomb interactions will be mainly determined not by pair collisions, but by the average forces acting on the particles. It is useful to note that inequality (32.1) is equivalent to the inequality

$$l_D^3 / \bar{r}^3 \gg 1. \quad (32.1')$$

This means that for the collision integral to be small it is necessary that the plasma parameter be small (see §41 of Part IV). This requirement is in its turn equivalent to the approximation of Debye's theory for an equilibrium plasma. Thus in the Debye approximation in which the equilibrium properties of a plasma are usually described it is possible to take into account collective interactions and to disregard pair interactions at small distances.

Inequality (32.1) or the equivalent inequality (32.1') is valid for low density and high temperature plasmas.

One can arrive at the same result by means of another argument based on the approximation of the relaxation time τ . If, in a plasma, one considers non-stationary processes with frequency ω , then for

$$\omega \gg 1/\tau$$

one can disregard the collision integral. Dropping the collision integral in the kinetic equation, one can write

$$\frac{\partial f_\alpha}{\partial t} + (\mathbf{v} \cdot \nabla) f_\alpha + \frac{\mathbf{F}_\alpha}{m} \cdot \nabla_{\mathbf{v}} f_\alpha = 0, \quad (32.2)$$

where the index α denotes the type of particle (electron, ion), and \mathbf{F}_α is the total force acting on a particle of type α

$$\mathbf{F}_\alpha = e_\alpha \left(\mathbf{E} + \frac{1}{c} [\mathbf{v} \times \mathbf{H}] \right). \quad (32.3)$$

Here we assume that $\mu \simeq 1$. The fields \mathbf{E} and \mathbf{H} in (32.3) represent the fields of all the other particles (internal fields) acting on the particle plus the applied external fields. For simplicity we shall first discuss the case where external fields are absent. The fields \mathbf{E} and \mathbf{H} satisfy the system of Maxwell-Lorentz equations

$$\begin{aligned}\nabla \times \mathbf{H} &= \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \mathbf{j}, \\ \nabla \times \mathbf{E} &= -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t}, \\ \nabla \cdot \mathbf{E} &= 4\pi\rho, \\ \nabla \cdot \mathbf{H} &= 0.\end{aligned}\tag{32.4}$$

We stress that in our approximation we cannot consider the plasma as a continuous medium and use Maxwell's equations.

The charge density and current density involved in the system of Maxwell-Lorentz equations (32.4) represent the mean values of these quantities taken with respect to the distribution functions f_α :

$$\rho = \sum_{\alpha} e_{\alpha} \int f_{\alpha}(\mathbf{r}, \mathbf{v}, t) dV, \quad \mathbf{j} = \sum_{\alpha} e_{\alpha} \int \mathbf{v} f_{\alpha}(\mathbf{r}, \mathbf{v}, t) dV. \tag{32.5}$$

Thus the set of field equations and distribution functions forms a closed system; the fields \mathbf{E} and \mathbf{H} are found from given mean values of the charge and current densities. The latter depend on the distribution function f_α . Thus it can be said that the fields \mathbf{E} and \mathbf{H} are defined by the distribution function f_α which in its turn, according to (32.2), depends on the values of the fields \mathbf{E} and \mathbf{H} at each point of the plasma at each instant of time. In other words, a spatial and velocity distribution of particles is established in the system such that the corresponding fields maintain this distribution. Each particle moves in a field produced by all the particles except the given one. Of course, in this respect all the particles are equivalent. Thus a self-consistent field is established in the plasma (see §41 of Part IV and §70 of Part V).

As a rule, one can disregard the thermal motion of ions because of their large mass and seek the distribution function just for the electrons. Then the whole system of ions forms a positively charged background compensating for the electric charge of the electrons.

Eq. (32.2) is called the kinetic equation with a self-consistent field. It should be noted that the qualitative reasoning underlying its derivation can be supplemented by more strict quantitative considerations. It then turns out that the kinetic equation with a self-consistent field is obtained from a set of equations for correlation functions by expanding in terms of a small plasma parameter r/l_D^* .

The collision integral turns out to be a quantity of the next order of small quantities in the expansion in terms of this parameter.

The self-consistent field approximation for non-equilibrium processes in a plasma allows one to discover a number of phenomena which occur in it.

§33. The dispersion and damping of plasma waves

We have already considered the theory of plasma waves in Part IV. We obtained the law of dispersion for plasma waves in the approximation where the plasma can be considered to be a continuous medium (i.e. in the hydrodynamic approximation). Here, however, we can considerably refine the macroscopic theory and, in particular, find the law of dispersion and damping of plasma waves.

Let us consider the high-frequency natural oscillations of a spatially uniform plasma in the absence of an external electromagnetic field. The kinetic equation for the electron distribution function in the self-consistent field approximation is of the form

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{e}{m} \left(\mathbf{E} + \frac{1}{c} [\mathbf{v} \times \mathbf{H}] \right) \cdot \frac{\partial f}{\partial \mathbf{v}} = 0. \quad (33.1)$$

Heavy ions are considered to be at rest. Assuming the field to be weak, we shall try to find the solution of the system of equations (33.1) and (32.4) in the form

$$f = f_0 + f_1, \quad (33.2)$$

where f_0 is the electron distribution function in the absence of oscillations,

* Note. See, for example, N.N.Bogoliubov, *Problems of dynamic theory in statistical physics*, in 'Studies in statistical mechanics', Ed. J.de Boer and G.E.Uhlenbeck, Vol. 1 (North-Holland, Amsterdam, 1962). K.P.Gurov, *Osnovaniya kineticheskoi teorii (Foundations of kinetic theory)* (Nauka, Moscow, 1966).

and $f_1 \ll f_0$. Substituting this expression into (33.1), we easily find for f_1

$$\frac{\partial f_1}{\partial t} + \mathbf{v} \cdot \frac{\partial f_1}{\partial \mathbf{r}} + \frac{e}{m} \left(\mathbf{E} + \frac{1}{c} [\mathbf{v} \cdot \mathbf{H}] \right) \cdot \frac{\partial f_0}{\partial \mathbf{v}} = 0. \quad (33.3)$$

Here f_1 is a function of coordinates, velocities and time, i.e.

$$f_1 = f_1(\mathbf{r}, \mathbf{v}, t).$$

The kinetic equation without the collision integral can be simplified considerably. In this approximation it can be assumed that all the particles move in defined trajectories under the action of the self-consistent and external fields. Hence, if \mathbf{r}_0 is the coordinate of the particle at the instant of time $t = 0$, then one can write

$$\mathbf{r} = \mathbf{r}_0 + \int_0^t \mathbf{v} dt, \quad (33.4)$$

where $\mathbf{v}(t)$ is defined by the equation of motion of the particle,

$$\mathbf{v}(t) = \mathbf{v}_0 + \int_0^t \frac{\mathbf{F}}{m} dt. \quad (33.5)$$

We introduce as new variables the values of \mathbf{r}_0 and \mathbf{v}_0 . Then

$$\begin{aligned} \left(\frac{\partial f_1}{\partial t} \right)_{\mathbf{r}_0, \mathbf{v}_0} &= \left(\frac{\partial f_1}{\partial t} \right)_{\mathbf{r}, \mathbf{v}} + \frac{\partial f_1}{\partial \mathbf{r}} \cdot \left(\frac{\partial \mathbf{r}}{\partial t} \right)_{\mathbf{r}_0, \mathbf{v}_0} + \frac{\partial f_1}{\partial \mathbf{v}} \cdot \left(\frac{\partial \mathbf{v}}{\partial t} \right)_{\mathbf{r}_0, \mathbf{v}_0} \simeq \\ &\simeq \left(\frac{\partial f_1}{\partial t} \right)_{\mathbf{r}, \mathbf{v}} + \mathbf{v} \cdot \frac{\partial f_1}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial f_1}{\partial \mathbf{v}}. \end{aligned}$$

Here we have assumed that the derivatives $(\partial \mathbf{r} / \partial t)_{\mathbf{r}_0, \mathbf{v}_0}$ and $(\partial \mathbf{v} / \partial t)_{\mathbf{r}_0, \mathbf{v}_0}$ are the same as the velocity and the force for unperturbed motion. This is accurate to within quantities of the second order of smallness. Since the quantities mentioned are multiplied by the derivatives of the perturbed distribution function f_1 , such a replacement is legitimate. Then the kinetic equation (33.3)

takes the very simple form

$$\frac{\partial f_1(\mathbf{r}_0, \mathbf{v}_0, t)}{\partial t} + \frac{e\mathbf{E}(\mathbf{r}_0, t)}{m} \cdot \frac{\partial f_0}{\partial \mathbf{v}(t)} = 0. \quad (33.6)$$

For an isotropic plasma f_0 does not depend on angles. Hence in the expression for the force only the electric force $e\mathbf{E}^*$ need be retained. This equation can be integrated directly. If it is assumed that $f_1(\mathbf{r}_0, t) \rightarrow 0$, for $t \rightarrow -\infty$, then, integrating (33.6), we have

$$f_1 = -\frac{e}{m} \int_{-\infty}^t \mathbf{E}(\mathbf{r}_0, t') \cdot \frac{\partial f_0}{\partial \mathbf{v}(t')} dt' = -\frac{e}{m} \int_{-\infty}^t \mathbf{E}(\mathbf{r} - \mathbf{v}t', t') \cdot \frac{\partial f_0}{\partial \mathbf{v}(t')} dt'. \quad (33.7)$$

We stress once more that the possibility of using eqs. (33.4) and (33.5) and of introducing new variables is associated with the absence of collisions. Collisions perturb the motion of particles and deprive one of the possibility of using the concept of motion along trajectories. Hence the solution (33.7) is valid only for a plasma without collisions.

Let us expand the field: \mathbf{E} in a Fourier integral. Then for f_1 we have

$$f_1 = -\frac{e}{m} \int d\mathbf{k} \int d\omega \int_{-\infty}^t \mathbf{E}(\mathbf{k}, \omega) e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} e^{-i\mathbf{k} \cdot \mathbf{v}(t-t')} \cdot \frac{\partial f_0}{\partial \mathbf{v}} dt'. \quad (33.8)$$

The correction to the distribution function due to an individual field harmonic is of the form

$$f_1^{(\mathbf{k}, \omega)} = -\frac{e}{m} \int_{-\infty}^t \mathbf{E}(\mathbf{k}, \omega) e^{i(\mathbf{k} \cdot \mathbf{v}t' - \omega(t-t'))} \cdot \frac{\partial f_0}{\partial \mathbf{v}(t')} dt'. \quad (33.9)$$

Knowing the correction to the distribution function, one can find the corresponding electric current \mathbf{j}

$$\begin{aligned} j_i &= en \int v_i(t) f_1^{(\mathbf{k}, \omega)} d\mathbf{v} = \\ &= -\frac{e^2 n}{m} E_i(\mathbf{k}, \omega) \int d\mathbf{v} \int_{-\infty}^t v_j(t) \frac{\partial f_0}{\partial v_j(t')} e^{-i[\mathbf{k} \cdot \mathbf{v}t' - \omega(t-t')]} dt'. \end{aligned} \quad (33.10)$$

* The product $[\mathbf{v} \times \mathbf{H}] \cdot \frac{\partial f_0}{\partial \mathbf{v}} = 2[\mathbf{v} \times \mathbf{H}] \cdot \mathbf{v} \frac{\partial f_0}{\partial v^2} = 0$.

We introduce the new variable $\tau = t - t'$. Then

$$j_i = -\frac{e^2 n}{m} E_i(\mathbf{k}, \omega) \int d\mathbf{v} \int_0^\infty v_j(0) \frac{\partial f_0}{\partial v_j(t-\tau)} e^{i(\mathbf{k} \cdot \mathbf{v}(t-\tau) + \omega\tau)} d\tau. \quad (33.11)$$

Since the time t is not specified, one can set $t = 0$ in the integral, and the current generated by one field harmonic turns out to be equal to

$$j_i = -\frac{e^2 n}{m} E_i(\mathbf{k}, \omega) \int d\mathbf{v} \int v_j(0) \frac{\partial f_0}{\partial v_j(\tau)} e^{i\omega\tau} e^{-i\mathbf{k} \cdot \mathbf{v}\tau} d\tau. \quad (33.12)$$

We stress that here f_0 denotes the electron distribution function in the plasma in the absence of oscillations. Generally speaking, f_0 is not an equilibrium distribution function.

If, however, the plasma is considered to be an equilibrium plasma, then f_0 represents a Maxwell distribution (normalized to unity), and

$$\frac{\partial f_0}{\partial v_j} = \frac{\partial f_M}{\partial v_j} = -\frac{mv_j}{kT} f_M. \quad (33.13)$$

Then formula (33.11) takes the form

$$j_i = \frac{e^2 n}{kT} E_i(\mathbf{k}, \omega) e^{i(\mathbf{k} \cdot \tau - \omega t)} \int d\mathbf{v} \int_0^\infty v_j(0) v_j(\tau) e^{i(\omega - \mathbf{k} \cdot \mathbf{v})\tau} f_M d\tau. \quad (33.14)$$

In §36 we shall discuss this expression which is of a general character and relates the current to the velocity correlation at instants of time separated by the interval τ . However, in the case considered, where external electric and magnetic fields are absent, the particles move with a constant velocity, so that $\mathbf{v}(0) = \mathbf{v}(\tau) = \text{const}$. Hence we shall drop the value of the argument of the velocity in subsequent formulae.

Formula (33.12) gives the following expression for the electrical conductivity tensor defined by the relation $j_i(\omega) = \sigma_{ik}(\omega) E_k(\omega)$

$$\sigma_{ij} = -\frac{e^2 n}{m} \int d\mathbf{v} \int v_i \frac{\partial f_0}{\partial v_j} e^{i(\omega - \mathbf{k} \cdot \mathbf{v})\tau} d\tau. \quad (33.15)$$

According to the general formula (33.20) of Part IV, the dielectric permeabil-

ity tensor ϵ_{ij} is of the form

$$\epsilon_{ij} = \delta_{ij} + i \frac{4\pi}{\omega} \sigma_{ij} = \delta_{ij} - i \frac{4\pi e^2 n}{m\omega} \int d\mathbf{v} \int v_i \frac{\partial f}{\partial v_j} e^{i(\omega - \mathbf{k} \cdot \mathbf{v})\tau} d\tau \quad (33.16)$$

or, for an equilibrium plasma

$$\begin{aligned} \epsilon_{ij} &= \delta_{ij} + i \frac{4\pi e^2 n}{\omega k T} \int f_M d\mathbf{v} \int v_i v_j e^{i(\omega - \mathbf{k} \cdot \mathbf{v})\tau} d\tau = \\ &= \delta_{ij} + i \frac{4\pi e^2 n}{\omega k T} \text{Av} \left\{ \int v_i v_j e^{i(\omega - \mathbf{k} \cdot \mathbf{v})\tau} d\tau \right\}, \end{aligned} \quad (33.17)$$

where $\text{Av} \{ \}$ denotes averaging over the Maxwell distribution.

We choose the direction of the vector \mathbf{k} to be the z -axis. It is then easily seen that in an isotropic plasma, which we are considering, only three components of the tensor ϵ_{ij} are different from zero: those perpendicular to the direction of propagation of the wave

$$\epsilon_{\perp} = \epsilon_{xx} = \epsilon_{yy} = 1 + i \frac{4\pi e^2 n}{\omega k T} \overline{(v_x^2)} \left\langle \int_0^{\infty} e^{i(\omega - kv_z)\tau} d\tau \right\rangle = 1 + i \frac{4\pi e^2 n}{\omega m} \langle \mathcal{U}_1 \rangle \quad (33.18)$$

and the component parallel to this direction

$$\epsilon_{\parallel} = \epsilon_{zz} = 1 + i \frac{4\pi e^2 n}{\omega k T} \left\langle v_z^2 \int_0^{\infty} e^{i(\omega - kv_z)\tau} d\tau \right\rangle = 1 + i \frac{4\pi e^2 n}{k T \omega} \langle \mathcal{U}_2 \rangle, \quad (33.19)$$

where $\langle \rangle$ denotes the average over the velocity component v_z , so that

$$\langle \mathcal{U}_1 \rangle = \left(\frac{m}{2\pi k T} \right)^{\frac{1}{2}} \int_{-\infty}^{\infty} \exp(-mv_z^2/2kT) \left(\int_0^{\infty} e^{i(\omega - kv_z)\tau} d\tau \right) dv_z, \quad (33.20)$$

$$\langle \mathcal{U}_2 \rangle = \left(\frac{m}{2\pi k T} \right)^{\frac{1}{2}} \int_{-\infty}^{\infty} v_z^2 \exp(-mv_z^2/2kT) \left(\int_0^{\infty} e^{i(\omega - kv_z)\tau} d\tau \right) dv_z. \quad (33.21)$$

Knowing the longitudinal and transverse components of the dielectric con-

stant tensor, one can find the law of dispersion of the longitudinal and transverse waves (see §33 of Part IV). Namely, the law of dispersion of the longitudinal waves is defined by the condition

$$\epsilon_{\parallel} = 0, \quad (33.22)$$

whereas the law of dispersion of the transverse waves is obtained from the relation

$$\epsilon_{\perp} - \frac{k^2 c^2}{\omega^2} = 1 + i \frac{4\pi e^2 n}{m\omega} \langle I_1 \rangle - \frac{k^2 c^2}{\omega^2} = 0, \quad (33.23)$$

We shall begin with a consideration of the dispersion of longitudinal waves. Taking into account (33.22) and (33.19), we find the transcendental equation connecting ω and k :

$$1 + i \frac{4\pi e^2 n}{kT\omega} \langle I_2 \rangle = 0. \quad (33.24)$$

Let us find, first of all, the integral with respect to the variable τ . For this integral we have a definite meaning, it must be considered as the limit

$$I_1 = \int_0^{\infty} \exp [i(\omega - kv_z)\tau] d\tau = \lim_{\gamma \rightarrow 0} \int_0^{\infty} e^{-\gamma\tau} \exp [i(\omega - kv_z)\tau] d\tau.$$

An arbitrarily small value of γ ensures the convergence of the integral at the upper limit. We then obtain

$$\int_0^{\infty} \exp [i(\omega - kv_z)\tau] d\tau = \lim_{\gamma \rightarrow 0} \frac{i}{(\omega - kv_z) + i\gamma}. \quad (33.25)$$

We now have to calculate the mean value of (33.25), i.e. the integral

$$\langle I_1 \rangle = \frac{i}{(\pi)^{\frac{1}{2}} kv_0} \int_{-\infty}^{\infty} e^{-(v/v_0)^2} \frac{d(v/v_0)}{\frac{i\gamma}{kv_0} + \left(\frac{\omega}{kv_0} - \frac{v}{v_0} \right)} = \frac{i}{(\pi kv_0)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \frac{e^{-x^2} dx}{z - x}, \quad (33.26)$$

where $z = (\omega + i\gamma)/kv_0$, $v_0 = (2kT/m)^{1/2}$. Instead of v_z we have written simply v . We make use of the identity valid for $\text{Im } z > 0$

$$\frac{i}{z-x} = \int_0^\infty e^{i(z-x)\alpha} d\alpha.$$

We then find

$$\begin{aligned} \langle I_1 \rangle &= \frac{1}{(\pi)^{1/2} kv_0} \int_{-\infty}^\infty e^{-x^2} \int_0^\infty e^{i(z-x)\alpha} dx d\alpha = \\ &= \frac{1}{(\pi)^{1/2} kv_0} \int_0^\infty e^{iz\alpha - \frac{1}{4}\alpha^2} d\alpha \int_{-\infty}^\infty e^{-(x+\frac{1}{2}\alpha)^2} dx = \\ &= \frac{1}{kv_0} \int_0^\infty e^{iz\alpha - \frac{1}{4}\alpha^2} d\alpha = \frac{e^{-z^2}}{kv_0} \int_0^\infty e^{-(\frac{1}{2}\alpha - iz)^2} d\alpha = \\ &= \frac{2e^{-z^2}}{kv_0} \left\{ \int_0^\infty e^{-t^2} dt - \int_0^{\bar{iz}} e^{-t^2} dt \right\} = \\ &= \frac{(\pi)^{1/2}}{kv_0} e^{-z^2} \left(1 + \frac{2i}{(\pi)^{1/2}} \int_0^z e^{w^2} dw \right) = \frac{(\pi)^{1/2}}{kv_0} e^{-z^2} + \frac{2i}{kv_0} e^{-z^2} \int_0^z e^{w^2} dw. \end{aligned}$$

In this last formula we can pass to the limit $\gamma \rightarrow 0$. We then finally obtain

$$\langle I_1 \rangle = \frac{(\pi)^{1/2}}{kv_0} e^{-(\omega/kv_0)^2} + \frac{2i}{kv_0} e^{-(\omega/kv_0)^2} \int_0^{\omega/kv_0} e^{w^2} dw. \quad (33.27)$$

It is obvious that $\omega/k = c_{ph}$ is the phase velocity of propagation of the waves, so that $z = \omega/kv_0$ represents the ratio of this velocity to the mean velocity of thermal motion of the electrons.

For $c_{ph} \gg v_0$ we can write

$$\begin{aligned} \int_0^z e^{w^2} dw &= e^{z^2} \int_0^z e^{-2zt} e^{t^2} dt = e^{z^2} \int_0^z e^{-2zt} \sum_{n=0}^{\infty} \frac{t^{2n}}{n!} dt \simeq \\ &\simeq e^{z^2} \sum_{n=0}^{\infty} \int_0^{\infty} e^{-2zt} \frac{t^{2n}}{n!} dt = \frac{e^{z^2}}{2z} \left(1 + \frac{1}{2z^2} + \dots \right), \end{aligned}$$

Hence

$$\langle J_1 \rangle \simeq \frac{(\pi)^{\frac{1}{2}}}{kv_0} e^{-(\omega/kv_0)^2} + \frac{i}{\omega} \left[1 + \frac{1}{2} \left(\frac{kv_0}{\omega} \right)^2 + \dots \right]. \quad (33.28)$$

To calculate the mean value in formula (33.21) we make use of the property of the integral $\langle J_1 \rangle$ which follows from its definition (33.25)

$$(\omega - kv_z) I_1 = i$$

or

$$v_z I_1 = \frac{\omega I_1}{k} - \frac{i}{k}.$$

Analogously

$$(\omega - kv_z)^2 I_1 = i(\omega - kv_z)$$

or

$$v_z^2 I_1 = \frac{\omega^2 I_1}{k^2} - \frac{i\omega}{k^2} - \frac{ikv_z}{k^2}. \quad (33.29)$$

Averaging (33.29), we find

$$\langle J_2 \rangle = \langle v_z^2 I_1 \rangle = \frac{\omega^2}{k^2} \langle J_1 \rangle - \frac{i\omega}{k^2}.$$

Substituting the value of $\langle I_1 \rangle$ from (33.28) for $c_{ph}' \gg v_0$, we finally obtain

$$\langle I_2 \rangle \simeq \frac{\omega^2}{k^2} \frac{(\pi)^{\frac{1}{2}}}{(kv_0)} e^{-(\omega/kv_0)^2} + \frac{i}{\omega} \left(\frac{\omega^2}{k^2} \right) \frac{(kv_0)^2}{2\omega^2}. \quad (33.30)$$

Substituting the value of $\langle I_2 \rangle$ into (33.24), we arrive at the following law of dispersion for longitudinal waves:

$$1 + i \frac{2(\pi)^{\frac{1}{2}} \omega_0^2}{\omega^2} \left(\frac{\omega}{kv_0} \right)^3 e^{-(\omega/\omega_0)^2} - \frac{\omega_0^2}{\omega^2} = 0, \quad (33.31)$$

where $\omega_0 = (4\pi e^2 n/m)^{\frac{1}{2}}$ is the plasma frequency (Langmuir frequency).

In the same approximation the solution of eq. (33.31) reads

$$\omega \simeq \omega_0(1 + i\delta), \quad (33.32)$$

where the decrement of damping δ is equal to

$$\delta = (\pi)^{\frac{1}{2}} \left(\frac{\omega_0}{kv_0} \right)^2 e^{-(\omega_0/kv_0)^2} = \frac{(\pi)^{\frac{1}{2}}}{2\sqrt{2}} \frac{1}{(kl_D)^3} e^{-\frac{1}{2}(1/kl_D)^2}, \quad (33.33)$$

Here l_D is the Debye length.

In §42 of Part IV, where a plasma was considered in the continuous-medium approximation, we established that limiting waves of frequency ω_0 and phase velocity $c_{ph} = \omega_0/k$ exist in the plasma. The damping of plasma waves is a new result. At first sight this result may seem paradoxical. We have seen before that dissipative processes are associated with molecular collisions and the resultant momentum transfer. The damping of plasma waves (the so-called Landau damping) found above is of a different nature.

Electrons in a plasma may have velocity components v_z both smaller and larger than the phase velocity of the wave, c_{ph} . In the first case the particles are acted upon by the field of a wave moving faster than themselves. Transferring momenta to the particles, the wave carries them along. On the contrary, those particles which are moving more rapidly than the wave lose their momentum, transferring it to the wave. Only particles whose velocity $v_z = c_{ph}$ are in resonance with the wave. They move in phase with the wave, neither losing nor gaining momentum. The longitudinal wave tends to distort the Maxwell distribution, producing a peak corresponding to a velocity v_z equal to the phase velocity c_{ph} .

It is known, however, that in an ensemble of particles having a Maxwell distribution the number of particles with velocity lower than a given velocity is larger than the number of particles with velocity larger than the given velocity. Hence the number of particles carried along by the wave exceeds the number of particles transferring momentum to the wave. As a result of this, a damping rather than an intensification of longitudinal waves takes place.

Formula (33.33) shows that the damping δ is small for wavelengths substantially exceeding the Debye length. On the contrary, the existence of plasma waves with $\lambda < l_D$ is impossible: the damping factor for such waves becomes larger than one.

This fact shows once more that longitudinal waves in a plasma represent a collective effect associated with the Coulomb interaction of charged particles.

Completely analogous calculations, based on the use of eq. (33.23) and of the expression for $\langle U_1 \rangle$ lead to the following expression for the frequency of transverse waves

$$\omega^2 = \omega_0^2 + k^2 c^2 > \omega_0^2. \quad (33.34)$$

Their phase velocity turns out to be larger than the velocity of light in vacuum

$$c_{ph} = \frac{\omega}{k} = c \left(1 + \frac{\omega_0^2}{c^2 k^2} \right)^{\frac{1}{2}}. \quad (33.35)$$

Hence in order to calculate the damping factor of transverse waves, one would have to take into account relativistic effects. However, it is clear that the damping effect is small, since the phase velocity of the waves is so large in comparison with the velocities of any electrons in the plasma that all the electrons can be considered to be at rest.

We shall not dwell here on the problem of the behaviour of a plasma in an external electromagnetic field, which is very important in practice. The method of integrating the kinetic equation in the presence of an external magnetic field does not differ in principle from that presented above, but requires more cumbersome calculations.

For this reason we do not consider the problem of plasma waves taking into account the motion of the heavy ions.

§34. The kinetic equation for a plasma taking into account collisions

The consideration of a plasma in the self-consistent field approximation turns out to be inadequate for describing a number of plasma processes.

Such processes are, first of all, relaxation processes (the establishment of a Maxwell distribution) as well as the levelling of the mean energies of electrons and ions. The self-consistent field approximation is also inadequate for calculating kinetic coefficients; coefficients of diffusion, of viscosity and so on.

For a sufficiently rarefied plasma one may take into account only pair collisions and make use of the ordinary Boltzmann kinetic equation but with a modified collision integral. This modification is associated with the properties of the Coulomb interaction. Since Coulomb interaction forces decrease slowly with increasing distance, the major contribution to the collision integral is given by collisions in which the particles are scattered at small angles. Indeed, for the Coulomb interaction the cross section diverges for small scattering angles (see §43 of Part I and §86 of Part V).

In small-angle scattering the change of momentum of the colliding particles is small. Hence it can be said that the main role is played by collisions in which a relatively small momentum transfer takes place. This can be used to transform the collision integral.

It is convenient to introduce new variables into the expression for the collision integral

$$I = \int w(\mathbf{p}_1, \mathbf{p}, \mathbf{p}_3, \mathbf{p}_2) [f(\mathbf{p}_3)f(\mathbf{p}_2) - f(\mathbf{p})f(\mathbf{p}_1)] d\mathbf{p}_1 d\Omega. \quad (34.1)$$

We assume that the change in the momentum of the first particle is

$$\mathbf{p}_2 - \mathbf{p} = \mathbf{q}. \quad (34.2)$$

From the momentum conservation law it follows that

$$\mathbf{p}_3 - \mathbf{p}_1 = -\mathbf{q}.$$

We carry out a change of variables, setting

$$\mathbf{p}_2 \rightarrow \frac{1}{2}(\mathbf{p}_2 + \mathbf{p}); \quad \mathbf{p}_3 \rightarrow \frac{1}{2}(\mathbf{p}_3 + \mathbf{p}_1);$$

$$\mathbf{p}_1 \rightarrow \mathbf{p}_3 - \mathbf{p}_1; \quad \mathbf{p} \rightarrow \mathbf{p}_2 - \mathbf{p}.$$

Then, obviously,

$$\begin{aligned} w(\mathbf{p}_3, \mathbf{p}_2, \mathbf{p}_1, \mathbf{p}) &\rightarrow w(\tfrac{1}{2}(\mathbf{p}_3 + \mathbf{p}_1), \tfrac{1}{2}(\mathbf{p}_2 + \mathbf{p}), \mathbf{p}_3 - \mathbf{p}_1, \mathbf{p}_2 - \mathbf{p}) \equiv \\ &\equiv w(\mathbf{p} + \tfrac{1}{2}\mathbf{q}, \mathbf{p}_1 - \tfrac{1}{2}\mathbf{q}, -\mathbf{q}, \mathbf{q}). \end{aligned}$$

Since the transition probability is invariant under the replacement $\mathbf{q} \rightarrow (-\mathbf{q})$ one can simply write by virtue of the principle of detailed balance,

$$w(\mathbf{p}_3, \mathbf{p}_2, \mathbf{p}_1, \mathbf{p}) \rightarrow w(\mathbf{p} + \tfrac{1}{2}\mathbf{q}, \mathbf{p}_1 - \tfrac{1}{2}\mathbf{q}, \mathbf{q}).$$

For such a change of variables we have

$$f(\mathbf{p}_3)f(\mathbf{p}_2) = f(\mathbf{p} + \mathbf{q})f(\mathbf{p}_1 - \mathbf{q}). \quad (34.3)$$

Then the collision integral assumes the form

$$I = \int w(\mathbf{p} + \tfrac{1}{2}\mathbf{q}, \mathbf{p}_1 - \tfrac{1}{2}\mathbf{q}, \mathbf{q}) [f(\mathbf{p} + \mathbf{q})f(\mathbf{p}_1 - \mathbf{q}) - f(\mathbf{p}_1)f(\mathbf{p})] d\mathbf{p}_1 d\Omega. \quad (34.4)$$

We now make use of the fact that the relative momentum change \mathbf{q} in a collision is small, and expand the distribution function f and the transition probability w in series in powers of q , retaining in the expansion terms no higher than the second order of small quantities. We then obtain

$$\begin{aligned} f(\mathbf{p} + \mathbf{q}) &\simeq f(\mathbf{p}) + \frac{\partial f}{\partial \mathbf{p}} \cdot \mathbf{q} + \frac{1}{2} \frac{\partial^2 f}{\partial p^i \partial p^k} q_i q_k, \\ f(\mathbf{p}_1 - \mathbf{q}) &\simeq f_1(\mathbf{p}_1) - \frac{\partial f_1}{\partial \mathbf{p}_1} \cdot \mathbf{q} + \frac{1}{2} \frac{\partial^2 f_1}{\partial p_1^i \partial p_1^k} q_i q_k, \\ [f(\mathbf{p} + \mathbf{q})f(\mathbf{p}_1 - \mathbf{q}) - f(\mathbf{p}_1)f(\mathbf{p})] &\simeq \left(f_1 \frac{\partial f}{\partial \mathbf{p}} - f \frac{\partial f_1}{\partial \mathbf{p}_1} \right) \cdot \mathbf{q} + \\ &+ \left(f_1 \frac{\partial^2 f}{\partial p^i \partial p^k} + f \frac{\partial^2 f_1}{\partial p_1^i \partial p_1^k} - 2 \frac{\partial f_1}{\partial p_1^i} \frac{\partial f}{\partial p^k} \right) \tfrac{1}{2} q_i q_k, \\ w(\mathbf{p} + \tfrac{1}{2}\mathbf{q}, \mathbf{p}_1 - \tfrac{1}{2}\mathbf{q}, \mathbf{q}) &\simeq w(\mathbf{p}, \mathbf{p}_1) + \left(\frac{\partial w}{\partial \mathbf{p}} - \frac{\partial w}{\partial \mathbf{p}_1} \right) \cdot \tfrac{1}{2} \mathbf{q} = \\ &= w + \left(\frac{\partial w}{\partial \mathbf{p}} - \frac{\partial w}{\partial \mathbf{p}_1} \right) \cdot \tfrac{1}{2} \mathbf{q}. \end{aligned}$$

Substituting these expressions into the collision integral, we find

$$\begin{aligned}
 I = & \int w \left(f_1 \frac{\partial f}{\partial \mathbf{p}} - f \frac{\partial f_1}{\partial \mathbf{p}_1} \right) \cdot \mathbf{q} \, d\mathbf{p}_1 \, d\Omega + \int \frac{\partial w}{\partial p^i} \left(f_1 \frac{\partial f}{\partial p^k} - f \frac{\partial f_1}{\partial p_1^k} \right) \frac{1}{2} q_i q_k \, d\mathbf{p}_1 \, d\Omega - \\
 & - \int \frac{\partial w}{\partial p_1^i} \left(f_1 \frac{\partial f}{\partial p^k} - f \frac{\partial f_1}{\partial p_1^k} \right) \frac{1}{2} q_i q_k \, d\mathbf{p}_1 \, d\Omega + \\
 & + \int w \left(f_1 \frac{\partial^2 f}{\partial p^i \partial p^k} + f \frac{\partial^2 f_1}{\partial p_1^i \partial p_1^k} - 2 \frac{\partial f_1}{\partial p_1^i} \frac{\partial f}{\partial p^k} \right) \frac{1}{2} q_i q_k \, d\mathbf{p}_1 \, d\Omega. \quad (34.5)
 \end{aligned}$$

By virtue of the evenness of the function w with respect to the change $\mathbf{q} \rightarrow -\mathbf{q}$ and the oddness of the integrand the first integral reduces to zero. The third integral can be transformed by integrating by parts:

$$\begin{aligned}
 & \int \frac{\partial w}{\partial p_1^i} \left(f_1 \frac{\partial f}{\partial p^k} - f \frac{\partial f_1}{\partial p_1^k} \right) \frac{1}{2} q_i q_k \, d\mathbf{p}_1 \, d\Omega = \\
 & = - \int w \left(\frac{\partial f_1}{\partial p_1^k} \frac{\partial f}{\partial p^k} - f \frac{\partial^2 f_1}{\partial p_1^i \partial p_1^k} \right) \frac{1}{2} q_i q_k \, d\mathbf{p}_1 \, d\Omega. \quad (34.6)
 \end{aligned}$$

Since the distribution function decreases rapidly with increasing argument, the integrated terms reduce to zero when $|\mathbf{p}_1| \rightarrow \infty$. Substituting (34.6) into (34.5), we find for the collision integral

$$\begin{aligned}
 I = & \int \left\{ \frac{\partial w}{\partial p^i} \left(f_1 \frac{\partial f}{\partial p^k} - f \frac{\partial f_1}{\partial p_1^k} \right) - w \left(\frac{\partial f_1}{\partial p_1^i} \frac{\partial f}{\partial p^k} - f \frac{\partial^2 f_1}{\partial p_1^i \partial p_1^k} \right) + \right. \\
 & \left. + w \left(f_1 \frac{\partial^2 f}{\partial p^i \partial p^k} + f \frac{\partial^2 f_1}{\partial p_1^i \partial p_1^k} - 2 \frac{\partial f_1}{\partial p_1^i} \frac{\partial f}{\partial p^k} \right) \right\} \frac{1}{2} q_i q_k \, d\mathbf{p}_1 \, d\Omega = \\
 & = \int \left\{ \frac{\partial w}{\partial p^i} \left(f_1 \frac{\partial f}{\partial p^k} - f \frac{\partial f_1}{\partial p_1^k} \right) - w \frac{\partial f_1}{\partial p_1^i} \frac{\partial f}{\partial p^k} + w f_1 \frac{\partial^2 f}{\partial p^i \partial p^k} \right\} \frac{1}{2} q_i q_k \, d\mathbf{p}_1 \, d\Omega = \\
 & = \int \left\{ \frac{\partial}{\partial p^i} \left[w \left(f_1 \frac{\partial f}{\partial p^k} - f \frac{\partial f_1}{\partial p_1^k} \right) \right] \right\} \frac{q_i q_k}{2} \, d\mathbf{p}_1 \, d\Omega = \\
 & = - \frac{\partial}{\partial p^i} \int w \left(f \frac{\partial f_1}{\partial p_1^k} - f_1 \frac{\partial f}{\partial p^k} \right) \frac{q_i q_k}{2} \, d\mathbf{p}_1 \, d\Omega = - \frac{\partial j_i}{\partial p^i}, \quad (34.7)
 \end{aligned}$$

where the following notation is introduced:

$$j_i = \int w \left(f \frac{\partial f_1}{\partial p^k} - f_1 \frac{\partial f}{\partial p^k} \right) \frac{1}{2} q_i q_k dp_1 d\Omega. \quad (34.8)$$

We have reduced the collision integral to the divergence of the vector j_i representing a flow in momentum space. The meaning of this result will become clear if the results of §10 are taken into account. When a variable, in the given case the momentum, changes by small amounts, then the change in the distribution function amounts to a flow in the corresponding space, in our case in momentum space.

The kinetic equation assumes the form

$$\frac{\partial(\mathbf{p}, \mathbf{r}, t)}{\partial t} + \mathbf{v} \cdot \frac{\partial f(\mathbf{p}, \mathbf{r}, t)}{\partial \mathbf{r}} + \frac{\mathbf{F}}{m} \cdot \frac{\partial f(\mathbf{p}, \mathbf{r}, t)}{\partial \mathbf{p}} = - \frac{\partial j_i}{\partial p^i}. \quad (34.9)$$

This equation is called Landau's kinetic equation. It is obvious that Landau's kinetic equation is a particular case of the equation of slow processes (the Fokker-Planck equation). In the case given, the slow process is the momentum exchange between particles via the Coulomb interaction.

A further simplification of the expression for momentum flow is obtained if integration over angles is carried out. For this we make use of the fact that the main role is played by distant collisions in which small-angle deflections occur.

We introduce the tensor

$$\alpha_{ik} = \int \frac{1}{2} w q_i q_k d\Omega. \quad (34.10)$$

By means of this tensor the expression for j_i can be written in the form

$$j_i = \int \alpha_{ik} \left(f \frac{\partial f_1}{\partial p^k} - f_1 \frac{\partial f}{\partial p^k} \right) dp_1. \quad (34.11)$$

In §43 of Part I we calculated the change in momentum for a small-angle deflection. If the direction of relative velocity of two colliding particles is chosen to be the x -axis, then we obviously have

$$q_x = 0; \quad q_y = q_z = \int \frac{\partial}{\partial y} \frac{e_1 e_2}{r} dt = \frac{2e_1 e_2}{\rho v_{\text{rel}}}, \quad (34.12)$$

where ρ is the impact parameter, and e_1, e_2 are the charges of the particles. In the general case the vector \mathbf{q} can be written in the form

$$\mathbf{q} = \frac{2e_1 e_2}{\rho^2 v_{\text{rel}}} \mathbf{p}. \quad (34.13)$$

Here it is obvious that the vector \mathbf{p} is perpendicular to the relative velocity vector \mathbf{v}_{rel} . Using (34.13), we write (34.11) as follows:

$$\alpha_{ik} = \frac{2(e_1 e_2)^2}{v_{\text{rel}}^2} \int \frac{w \rho_i \rho_k}{\rho^4} d\Omega. \quad (34.14)$$

The probability w of a collision with scattering into angle $d\Omega$ can be written in the form

$$w d\Omega = dw = v_{\text{rel}} \rho d\varphi, \quad (34.15)$$

where φ is the azimuthal scattering angle (the angle specifying the direction of \mathbf{p} in a plane perpendicular to the vector \mathbf{v}_{rel} containing \mathbf{p}).

Taking into account this value of w and (34.14), we obtain

$$\alpha_{ik} = \frac{2(e_1 e_2)^2}{v_{\text{rel}}} \int \frac{\rho_i \rho_k}{\rho^3} d\varphi d\rho.$$

In the coordinate system which we have chosen it is obvious that one can write

$$\rho_x = 0; \quad \rho_y = \rho \sin \varphi; \quad \rho_z = \rho \cos \varphi.$$

so that for the components α_{ik} we finally get

$$\alpha_{xx} = \alpha_{xy} = \alpha_{xz} = \alpha_{yz} = 0, \quad \alpha_{yy} = \alpha_{zz} = \frac{2\pi(e_1 e_2)^2}{v_{\text{rel}}} \int \frac{d\rho}{\rho}. \quad (34.16)$$

The integral over impact parameters diverges logarithmically at the upper as well as the lower limit. The limits of integration (the values of the parameters ρ_{max} and ρ_{min}) can be determined from the following consideration. When ρ exceeds the Debye length l_D , charged particles are screened and essentially do not interact. Hence the upper limit can be set as $\rho_{\text{max}} \simeq l_D$.

The lower limit is determined from the condition that scattering angles be not too large. Namely, if the kinetic energy $\frac{1}{2}\mu v_{\text{rel}}^2$ is large in comparison with the potential energy $e_1 e_2 / \rho$, then the deflections are relatively small.

The limit of the region of small deflections is defined by the condition

$$\frac{1}{2}\mu v_{\text{rel}}^2 \sim e_1 e_2 / \rho_{\text{min}}$$

or

$$\rho_{\text{min}} \sim 2e_1 e_2 / \mu v_{\text{rel}}^2.$$

Thus, finally,

$$\alpha_{ik} = \alpha_{zz} = \frac{2\pi(e_1 e_2)^2}{v_{\text{rel}}} \ln \frac{\mu v_{\text{rel}}^2 l_D}{2e_1 e_2}.$$

It should be noted that the quantity whose logarithm is to be taken is very large, so that the numerical value of the logarithm itself is not very sensitive to the definition of the parameters ρ_{max} and ρ_{min} .

In an arbitrary coordinate system, α_{ik} is written in the form

$$\alpha_{ik} = 2\pi(e_1 e_2)^2 \frac{v_{\text{rel}}^2 \delta_{ik} - v_{\text{rel}}^i v_{\text{rel}}^k}{v_{\text{rel}}^3} \ln \frac{\rho_{\text{max}}}{\rho_{\text{min}}}. \quad (34.17)$$

We finally obtain

$$\begin{aligned} \frac{\partial f}{\partial t} + v_k \frac{\partial f}{\partial r_k} + \frac{F_k}{m} \frac{\partial f}{\partial p_k} = \\ = -2\pi(e_1 e_2)^2 \ln \frac{\rho_{\text{max}}}{\rho_{\text{min}}} \frac{\partial}{\partial \rho_i} \int \frac{v_{\text{rel}}^2 \delta_{ik} - v_{\text{rel}}^i v_{\text{rel}}^k}{v_{\text{rel}}^3} \left(f \frac{\partial f_1}{\partial p_1^k} - f_1 \frac{\partial f}{\partial p^k} \right) dp_1. \end{aligned} \quad (34.18)$$

In plasmas which consist of particles of several kinds, the Landau equations should be written for the distribution function of the particles of each kind. We note, first of all, that for an equilibrium state the Landau equation allows a solution in the form of a Maxwell distribution

$$f = \text{const} e^{-p^2/2mkT}. \quad (34.19)$$

Indeed, on substituting (34.19) into the Landau equation its left-hand side reduces to zero. On the right-hand side we have

$$\begin{aligned} \int \frac{v_{\text{rel}}^2 \delta_{ik} - v_{\text{rel}}^i v_{\text{rel}}^k}{v_{\text{rel}}^3} \left(f \frac{\partial f_1}{\partial p_1^k} - f_1 \frac{\partial f}{\partial p^k} \right) dp_1 = \\ = \int \frac{v_{\text{rel}}^2 \delta_{ik} - v_{\text{rel}}^i v_{\text{rel}}^k}{v_{\text{rel}}^3} \left(\frac{p_1^k - p^k}{mkT} \right) ff_1 dp_1 = \frac{1}{kT} \int \frac{v_{\text{rel}}^k \delta_{ik} - v_{\text{rel}}^i}{v_{\text{rel}}} ff_1 dp_1. \end{aligned}$$

However, it is clear that

$$v_{\text{rel}}^k \delta_{ik} - v_{\text{rel}}^i = 0.$$

Thus the solution of the Landau equation for the case of equilibrium is a Maxwell distribution. It can be shown that the H -theorem follows from Landau's equation. Hence Maxwell's distribution is the only distribution established in an equilibrium plasma. Finally, the equations of plasma hydrodynamics follow from Landau's equation, and the corresponding kinetic coefficients can be found from it. However, we cannot dwell on these rather cumbersome calculations.

§35. The establishment of equilibrium in an electron-ion plasma

Owing to the substantial difference between the masses of ions and electrons, an electron-ion plasma represents a classical example of a system which can be in a state of incomplete equilibrium.

Collisions between ions lead to the establishment of an equilibrium Maxwell distribution with a certain temperature $T^{(i)}$ in time τ_i . Analogously, in time τ_e an equilibrium electron distribution will be established which, according to what was said in the preceding section, is also a Maxwell distribution. However, the value of electron temperature $T^{(e)}$ will differ from that of ion temperature $T^{(i)}$, namely $T^{(e)} \gg T^{(i)}$.

A plasma with different ion and electron temperatures is in a state of incomplete equilibrium. After the lapse of a time $\tau \gg \tau_i, \tau_e$, total equilibrium with a common temperature for the two kinds of particles will be established in the plasma. This equilibrium is established by means of energy transfer

from the electrons to the ions. The relaxation time τ is defined by the equality

$$\frac{d\overline{\epsilon^{(i)}}}{dt} = Q^{(e \rightarrow i)}, \quad (35.1)$$

where $\overline{\epsilon^{(i)}} = \frac{3}{2} k T_i$ is the mean ion energy, and $Q^{(e \rightarrow i)}$ is the energy flow from electrons to ions.

Obviously, we have

$$\frac{d\overline{\epsilon^{(i)}}}{dt} = \frac{d}{dt} \int \epsilon^{(i)} f^{(i)} d\mathbf{p}^{(i)} = \int \epsilon^{(i)} \frac{\partial f^{(i)}}{\partial t} d\mathbf{p}^{(i)},$$

since the ion distribution function $f^{(i)}$ does not depend on coordinates. According to (34.9) the distribution function satisfies the kinetic equation

$$\frac{\partial f^{(i)}}{\partial t} = - \frac{\partial}{\partial p_k^{(i)}} j_k^{(e, i)}. \quad (35.2)$$

The collision integral $j_k^{(i, i)} = 0$, since there is equilibrium between the ions. Substituting (35.2) into (35.1) and integrating by parts, we have

$$\frac{d\overline{\epsilon^{(i)}}}{dt} = - \int \epsilon^{(i)} \frac{\partial j_k^{(e, i)}}{\partial p_k^{(i)}} d\mathbf{p}^{(i)} = \int v_k^{(i)} j_k^{(e, i)} d\mathbf{p}^{(i)}. \quad (35.3)$$

According to (34.11), the momentum flow from the electron component to the ion component can be written in the form

$$j_k^{(e, i)} = \int \alpha_{kj} \left(f_0^{(e)} \frac{\partial f_0^{(i)}}{\partial p_j^{(i)}} - f_0^{(i)} \frac{\partial f_0^{(e)}}{\partial p_j^{(e)}} \right) d\mathbf{p}^{(e)}. \quad (35.4)$$

Since $f_0^{(e)}$ and $f_0^{(i)}$ are equilibrium distribution functions with temperatures $T^{(e)}$ and $T^{(i)}$, we find

$$j_k^{(e, i)} = \int \alpha_{jk} f_0^{(e)} f_0^{(i)} \left(\frac{v_j^{(i)}}{k T^{(i)}} - \frac{v_j^{(e)}}{k T^{(e)}} \right) d\mathbf{p}^{(e)}. \quad (35.5)$$

We introduce the relative velocity

$$v_{\text{rel}}^j = v_j^{(e)} - v_j^{(i)}.$$

Then

$$j_k^{(e,i)} = \int \alpha_{ik} f_0^{(e)} f_0^{(i)} \left\{ v_j^{(i)} \left(\frac{1}{kT^{(i)}} - \frac{1}{kT^{(e)}} \right) - \frac{v_{\text{rel}}^i}{kT^{(e)}} \right\} d\mathbf{p}^{(e)}.$$

However, we have

$$\alpha_{kj} v_{\text{rel}}^j \simeq v_{\text{rel}}^j v_{\text{rel}}^2 \delta_{kj} - v_{\text{rel}}^2 v_{\text{rel}}^k = 0.$$

Hence we find

$$j_k^{(e,i)} = \frac{2\pi e^4 Z^2}{k} \ln \frac{\rho_{\text{max}}}{\rho_{\text{min}}} \left(\frac{1}{T^{(i)}} - \frac{1}{T^{(e)}} \right) \int f_0^{(i)} f_0^{(e)} v_j^{(i)} \frac{v_{\text{rel}}^2 \delta_{kj} - v_{\text{rel}}^k v_{\text{rel}}^j}{v_{\text{rel}}^3} d\mathbf{p}^{(e)}.$$

Since the velocity of the electrons is large in comparison with the velocity of the ions, we can set

$$v_{\text{rel}} \simeq v^{(e)}.$$

And finally,

$$j_k^{(e,i)} = \frac{2\pi e^4 Z^2}{k} \frac{T^{(e)} - T^{(i)}}{T^{(i)} T^{(e)}} \ln \frac{\rho_{\text{max}}}{\rho_{\text{min}}} \int f_0^{(i)} f_0^{(e)} \frac{\delta_{kj} (v^{(e)})^2 - v_k^{(e)} v_j^{(e)}}{(v^{(e)})^3} v_j^{(i)} d\mathbf{p}^{(e)}. \quad (35.6)$$

Substituting $j_k^{(e,i)}$ from (35.6) into (35.3), we obtain

$$\begin{aligned} \frac{d\epsilon^{(i)}}{dt} &= \int v_k^{(i)} j_k^{(e,i)} d\mathbf{p}^{(i)} = \\ &= 2\pi \frac{e^4 Z^2}{k} \frac{T^{(e)} - T^{(i)}}{T^{(i)} T^{(e)}} \ln \frac{\rho_{\text{max}}}{\rho_{\text{min}}} \int f_0^{(e)} f_0^{(i)} v_k^{(i)} v_j^{(i)} \frac{(v^{(e)})^2 \delta_{kj} - v_k^{(e)} v_j^{(e)}}{(v^{(e)})^3} d\mathbf{p}^{(e)} d\mathbf{p}^{(i)}. \end{aligned} \quad (35.7)$$

Obviously, we have

$$v_k^{(i)} v_j^{(i)} (v^{(e)})^2 \delta_{kj} = (v^{(i)})^2 (v^{(e)})^2.$$

Furthermore, in view of the spherical symmetry of $f_0^{(i)}$ and $f_0^{(e)}$

$$\langle v_k^{(e)} v_j^{(e)} \rangle \langle v_k^{(i)} v_j^{(i)} \rangle = 0 \quad \text{for } k \neq j.$$

Hence, finally,

$$\frac{d\bar{\epsilon}^{(i)}}{dt} = 2\pi \frac{e^4 Z^2}{k} \frac{T^{(e)} - T^{(i)}}{T^{(i)} T^{(e)}} \ln \frac{\rho_{\max}}{\rho_{\min}} \int \frac{(v^{(i)})^2}{v^{(e)}} f_0^{(e)} f_0^{(i)} d\mathbf{p}^{(e)} d\mathbf{p}^{(i)}.$$

Subsequent calculations reduce to a simple averaging over the Maxwell ion and electron distributions:

$$\langle (v^{(i)})^2 \rangle = \frac{3kT^{(i)}}{m^{(i)}}, \quad \left\langle \frac{1}{v^{(e)}} \right\rangle = 2 \left(\frac{m^{(e)}}{2\pi kT^{(e)}} \right)^{\frac{1}{2}}.$$

Thus for $d\bar{\epsilon}^{(i)}/dt$ we get

$$\frac{d\bar{\epsilon}^{(i)}}{dt} = 2\pi \frac{N^2 e^4 Z^2}{k} \ln \frac{\rho_{\max}}{\rho_{\min}} \frac{T^{(e)} - T^{(i)}}{T^{(i)} T^{(e)}} \frac{3kT^{(i)}}{m^{(i)}} 2 \left(\frac{m^{(e)}}{2\pi kT^{(e)}} \right)^{\frac{1}{2}}$$

or

$$\frac{dT^{(i)}}{dt} = \frac{T^{(e)} - T^{(i)}}{\tau},$$

where the relaxation time is

$$\tau = \frac{3m^{(i)}}{4(2\pi)^{\frac{1}{2}} (m^{(e)})^{\frac{1}{2}}} \frac{(kT^{(e)})^{\frac{3}{2}}}{NZ^2 e^4 \ln(\rho_{\max}/\rho_{\min})}.$$

It is useful to compare this time with the relaxation time of each of the plasma components. The calculation of the latter involves great difficulties, but a rough estimate can be found from the following considerations. Setting $f = f_0 + f'$ in the kinetic equation, we have

$$\frac{\partial f'}{\partial t} = \frac{f - f_0}{\tau_{(e,i)}} = \frac{f'}{\tau_{(e,i)}}$$

where in order of magnitude

$$\frac{1}{\tau_{(e,i)}} \sim \frac{1}{m^{\frac{1}{2}}(kT)^{\frac{3}{2}}} \sim \int \frac{\partial^2 f_0}{\partial p^2} \frac{1}{v_{\text{rel}}} dp.$$

Thus

$$\tau^{(e)} \sim m_e^{\frac{1}{2}}(kT^{(e)})^{\frac{3}{2}},$$

$$\tau^{(i)} \sim m_i^{\frac{1}{2}}(kT^{(i)})^{\frac{3}{2}}.$$

The ratio of the electron relaxation time to the ion relaxation time is

$$\frac{\tau^{(e)}}{\tau^{(i)}} \sim \frac{m^{(e)}}{m^{(i)}}$$

i.e. is indeed very small.

The Time Correlation Function Method and Onsager's Theory

§36. The response of a system to a dynamic perturbation.

Classical calculation

So far we have discussed only one approach to solving kinetic problems; the method of the kinetic equation. The complexity of the basic kinetic equation makes it necessary to pass over to Boltzmann's equation for the one-particle distribution function for the actual solving of kinetic problems.

As we have seen in the examples given and as will be particularly clear in the chapter devoted to solid-state theory, Boltzmann's equation is a powerful method of investigating non-equilibrium processes. However, it allows one to obtain concrete results only for a limited class of systems.

Another method of solving kinetic problems has recently been developed. In this method it has been possible to formulate physical kinetics in the same way as statistical physics.

Let us consider a macroscopic system in a state of statistical equilibrium. The equilibrium properties of this system are described by the equilibrium density matrix or, in the classical approximation, by the Gibbs distribution. We now suppose that a small perturbation is switched on at the instant of time $t \rightarrow -\infty$.

In principle there are two different classes of perturbation. The first of these is associated with the application of an external force field to the sys-

tem; for example a time-dependent electric or magnetic field. We shall call such perturbations dynamic. When a dynamic perturbation is applied, the total Hamiltonian can be written in the form

$$H = H_0 + H'(t),$$

where $H'(t)$ describes that part of the Hamiltonian which is associated with the action of the external force. Thus dynamic perturbations are of a microscopic nature. They change the Hamiltonian of each particle of the system.

Perturbations of another class, often called thermal perturbations, are of macroscopic character and make sense only with respect to the system as a whole or to a macroscopic part of it. For example, when thermal or diffusion contact is established between bodies having different temperatures or different compositions, the states of each of the bodies undergo perturbation. However, such a perturbation cannot be related to the change of the Hamiltonian of the individual particles.

We shall subsequently discuss the action of dynamic and thermal perturbations on an equilibrium macroscopic system, restricting ourselves first to the quasi-classical approximation and then carrying out the quantum calculation.

It turns out that all kinetic coefficients, and consequently, also all transport coefficients, can be expressed in terms of one and the same quantity, the time correlation function. Thus in physical kinetics the time correlation function plays the same role as that played by the partition function in statistical physics.

Therefore let us consider a classical quasi-closed system characterized by the Gibbs distribution function

$$\rho_0 = \frac{1}{Z} e^{-\beta H_0} \quad (36.1)$$

in an equilibrium state.

The perturbed Hamiltonian function is written in the form

$$H = H_0 + H'(t),$$

where $H'(t) \ll H_0$. To simplify the calculations, we set

$$H' = -A[p(t), q(t)] \delta(t).$$

This means that at the instant of time $t = 0$ the system is acted upon by a certain impulse, whereas at $t < 0$ and $t > 0$ the system is not influenced by any external action. We shall find the changes caused in the system by a small dynamic perturbation.

The variation of the distribution function in time is given by the general formula (15.2)

$$\frac{\partial \rho}{\partial t} = \{H; \rho\}. \quad (36.2)$$

Assuming that the perturbed distribution function can be written in the form

$$\rho = \rho_0 + \rho', \quad (36.3)$$

where ρ' is a small addition due to the perturbation ($\rho' \ll \rho_0$), we find

$$\frac{\partial \rho'}{\partial t} = \{H'; \rho_0\} + \{H_0; \rho'\} = -\{A(p, q) \cdot \delta(t); \rho_0\} + \{H_0; \rho'\} \quad (36.4)$$

or

$$\frac{d\rho'}{dt} = -\{A(p, q); \rho_0\} \delta(t).$$

Integrating with respect to time, we obtain

$$\rho' = \rho'(-\infty) - \{A(0); \rho_0\} = -\{A(0); \rho_0\}, \quad (36.5)$$

where $A(0)$ denotes the expression $A[p(t), q(t)]$ in which the values of the coordinates and momenta of the particles at the instant of time $t = 0$ are taken.

By definition of the Poisson bracket and by virtue of (36.1), it is easy to find

$$\{A(0); \rho_0\} = \sum \left(\frac{\partial A}{\partial q} \frac{\partial \rho_0}{\partial p} - \frac{\partial A}{\partial p} \frac{\partial \rho_0}{\partial q} \right) = -\frac{1}{kT} \rho_0 \sum \left(\frac{\partial A}{\partial q} \frac{\partial H_0}{\partial p} - \frac{\partial A}{\partial p} \frac{\partial H_0}{\partial q} \right).$$

But for any mechanical quantity depending on coordinates and momenta one

can write

$$\dot{A} = \frac{dA}{dt} = \sum \left(\frac{\partial A}{\partial q} \frac{dq}{dt} + \frac{\partial A}{\partial p} \frac{dp}{dt} \right) = \sum \left(\frac{\partial A}{\partial q} \frac{\partial H_0}{\partial p} - \frac{\partial A}{\partial p} \frac{\partial H_0}{\partial q} \right).$$

Hence we obtain finally

$$\rho' = \dot{A}(0) \rho_0 / kT. \quad (36.6)$$

Knowing the change in the distribution function, one can find the change in the mean value of any quantity, $B(t)$, describing a macroscopic subsystem, caused by the perturbation

$$\begin{aligned} \langle \Delta B(t) \rangle &= \int [B(t)(\rho_0 + \rho') - B(t)\rho_0] d\Gamma = \\ &= \frac{1}{kT} \int \dot{A}(0) B(t) \rho_0 d\Gamma = \frac{1}{kT} \langle \dot{A}(0) B(t) \rangle. \end{aligned} \quad (36.7)$$

Formula (36.7) defines the change in the mean value of an arbitrary quantity B under the action of unit impulse. We denote this change by φ_{BA} and call it the response of the system

$$\varphi_{BA}(t) = \frac{1}{kT} \langle \dot{A}(0) B(t) \rangle. \quad (36.8)$$

Then (36.7) can be written in the form

$$\langle \Delta B \rangle = \int \varphi_{BA}(t - t') \delta(t') dt'. \quad (36.9)$$

Let us now consider the very general case where the perturbation acting on the system is of the form

$$H' = -A[p(t), q(t)] F(t), \quad (36.10)$$

where $F(t)$ is a certain given function of time. In the linear approximation, i.e. when the perturbation is small, the change in mean values under the action of the perturbation (36.10) can be considered as a superimposition of impulse perturbations, and instead of (36.9) one can write

$$\langle \Delta B \rangle = \int_{-\infty}^t \varphi_{BA}(t - t') F(t') dt', \quad (36.11)$$

where the response φ_{BA} , as before, is given by formula (36.8), so that

$$\langle \Delta B \rangle = \frac{1}{kT} \int_{-\infty}^t \langle \dot{A}(0) B(t-t') \rangle F(t') dt'. \quad (36.12)$$

However, it is necessary that $F(t)$ satisfy the very general requirement

$$F(t \rightarrow -\infty) \rightarrow 0. \quad (36.13)$$

which means that before switching on the perturbation the system was in an equilibrium state.

An important particular case is that of a periodic perturbation. In order to satisfy condition (36.13), it can be assumed that

$$F(t) = \text{Re} \lim_{\delta \rightarrow 0} e^{\delta t + i\omega t}. \quad (36.14)$$

Formula (36.14) defines a function which is practically periodic for all values of time t except for $t \rightarrow -\infty$ when it reduces to zero. Then from (36.11) we obtain

$$\langle \Delta B \rangle = \text{Re} \chi_{BA} e^{i\omega t}, \quad (36.15)$$

where

$$\chi_{BA} = \lim_{\delta \rightarrow 0} \int_0^{\infty} e^{-i\omega t' + \delta t'} \varphi_{BA} dt'. \quad (36.16)$$

This last quantity, representing the Fourier component of the response φ_{BA} , is called the generalized complex susceptibility. We shall see below that in the case of the action of an electromagnetic field, this definition of susceptibility is the same as that given in electrodynamics. The relations obtained have the same degree of generality as the relations of classical statistical physics. For small departures of the quasi-closed system from an equilibrium state, formula (36.6) defines the change in its distribution function, and formula (36.12) defines the corresponding change in the mean values.

Relations of the type (36.11) are accurate in the sense that they do not depend on the actual physical properties of the non-equilibrium system con-

sidered. Thus formula (36.12) defines the deviation of the mean values of quantities characterizing the system from their equilibrium values when the system is acted upon by dynamic perturbations. It turns out that the quantity characterizing the response of the system to a dynamic perturbation is the correlation function $\varphi_{BA} = (1/kT) \langle \dot{A}(0) B(t) \rangle$. Since the averaging in φ_{BA} is carried out over the states of the equilibrium system, formula (36.12) allows one to find the mean values for the non-equilibrium system from the characteristics of the equilibrium system. The correlation function or, more precisely, the response function for the non-equilibrium systems considered plays the same role as the distribution function for the equilibrium systems.

However, it should be kept in mind that the distribution function is of universal character. On the contrary, the response function depends on the nature of the perturbation (the quantity $\dot{A}(0)$).

Before going on to obtain formula (36.12) in quantum statistics and to a discussion of the applications of the general theory, we shall make one more remark on the essence of formula (36.12).

Formula (36.12), as well as formulae for the calculation of mean values in statistics, makes sense only for sufficiently large systems ($N \rightarrow \infty$, $V \rightarrow \infty$, finite N/V). Furthermore, it is necessary to consider the time t to be arbitrarily large, i.e. to know the response of the system to a perturbation a sufficiently long time after switching on the perturbation.

The function φ_{BA} is invariant under the replacement $t \rightarrow -t$:

$$\varphi_{BA} = \frac{1}{kT} \langle \dot{A}(0) B(t) \rangle = \frac{1}{kT} \langle \dot{A}(0) B(-t) \rangle,$$

since $B(t) = B(-t)$. On the contrary, $\langle \Delta B \rangle$ for sufficiently long times is not invariant under this replacement. This is obvious from simple physical reasoning. If t^* is the time lapse after switching on the perturbation, and $t > t^*$, then $\langle B(-t) \rangle$ is the response of the system to a perturbation which has not yet acted on it! Thus the expression for $\langle \Delta B \rangle$ turns out to be irreversible for sufficiently long times and for sufficiently large systems.

The relations obtained can be rewritten in a more convenient form, if use is made of the condition

$$\frac{d}{dt} \langle A(0) B(t) \rangle = \left\langle \frac{d}{dt} (A(0) B(0)) \right\rangle = 0. \quad (36.17)$$

which means that the mean equilibrium rate of change of correlation of two arbitrary quantities is equal to zero. This condition expresses the stationary character of the processes considered.

Taking into account (36.17), one can write

$$\varphi_{BA} = -\frac{1}{kT} \langle A(0) \dot{B}(t) \rangle \quad (36.18)$$

and, correspondingly,

$$\langle \Delta B \rangle = -\frac{1}{kT} \int_{-\infty}^t \langle A(0) \dot{B}(t-t') \rangle F(t') dt' . \quad (36.19)$$

We Fourier transform formula (36.19). We write

$$\begin{aligned} F(\omega) &= \int_0^{\infty} F(t) e^{-i\omega t} dt , \\ \langle J(\omega) \rangle &= \int_0^{\infty} \langle \Delta B \rangle e^{-i\omega t} dt = \\ &= -\frac{1}{kT} \int_{-\infty}^t dt' F(t') \int_0^{\infty} \langle A(0) \dot{B}(t-t') \rangle e^{-i\omega t} dt = \\ &= -\frac{1}{kT} \int_0^{\infty} \int_0^{\infty} du \langle A(0) \dot{B}(u) \rangle F(t-u) e^{-i\omega(t-u)} e^{-i\omega u} dt . \end{aligned}$$

Then (36.19) can be written as follows

$$\langle J(\omega) \rangle = \gamma(\omega) F(\omega) , \quad (36.20)$$

where $\gamma(\omega)$ is of the form

$$\gamma(\omega) = -\frac{1}{kT} \int_0^{\infty} e^{-i\omega u} \langle A(0) \dot{B}(u) \rangle du . \quad (36.21)$$

Formula (36.20) defines the generalized transport coefficient relating the

Fourier components of the 'force' F to the change in the quantity B caused by it.

It is natural to call $I(\omega)$ the generalized flux corresponding to the force $F(\omega)$. The mean flux $\langle I(\omega) \rangle$ turns out to be connected with the 'force' $F(\omega)$ by a relation which is a generalization of empirical relations well-known from electrodynamics, i.e. Ohm's law, relations for the electric and magnetic susceptibility, and so on.

Formula (36.21) can be generalized to the case of vector forces or several forces acting on the system. Instead of (36.21), one can at once write the equality

$$\langle J_k(\omega) \rangle = \gamma_{ik}(\omega) F_i(\omega),$$

where i, k run over the corresponding sequence of values (for example $k = x, y, z$ or $k = 1, 2, 3, \dots$). The tensor γ_{ik} is called the tensor of transport coefficients

$$\gamma_{ik}(\omega) = -\frac{1}{kT} \int_0^\infty e^{-i\omega u} \langle \dot{A}_i(0) \dot{B}_k(u) \rangle du.$$

§37. The response of a system to an external dynamic perturbation. Quantum calculation

Let us now consider the response of a system to an external dynamic perturbation by means of the quantum equation for the density matrix (2.3).

Let a quantum-mechanical system described by the density matrix $\hat{\rho}$ be in a reservoir and subjected to the action of an arbitrary external field $U(t)$ depending on time. We shall assume the external field to be sufficiently weak to be considered a small perturbation. Furthermore, we shall assume that the applied field satisfies the requirement

$$U(t) \rightarrow 0 \quad \text{for} \quad t \rightarrow -\infty. \quad (37.1)$$

We write, first of all, the equation for the statistical operator in the absence of an external field

$$i\hbar \frac{\partial \hat{\rho}}{\partial t} + [\hat{\rho}, \hat{H}] = 0. \quad (37.2)$$

When an external field $U(t)$ is applied, the equation for the statistical operator assumes the form

$$i\hbar \frac{\partial \hat{\rho}}{\partial t} + [\hat{\rho}, \hat{H} + U(t)] = 0. \quad (37.3)$$

It is convenient to pass from eq. (37.3) to an integral equation. For this we shall consider the term $[\rho, U]$ to be a known quantity. Then formally (37.3) will represent a linear non-homogeneous equation of first order with respect to the function ρ . If this equation is supplemented by the initial condition

$$\hat{\rho}(t \rightarrow \infty) \rightarrow \hat{\rho}_0, \quad (37.4)$$

then the solution can at once be written in the form

$$\hat{\rho} = \hat{\rho}_0 + \frac{i}{\hbar} \int_{-\infty}^t e^{-iH(t-x)/\hbar} [\hat{\rho}(x), U(x)] e^{iH(t-x)/\hbar} dx. \quad (37.5)$$

By direct substitution one can see that (37.5) satisfies eq. (37.3) and the initial condition (37.4). The integral equation (37.5) contains the small perturbation $U(x)$ and can be solved by iteration (successive approximations). Setting

$$\hat{\rho} = \hat{\rho}_0 + \hat{\rho}' + \hat{\rho}'' + \dots, \quad (37.6)$$

we have

$$\begin{aligned} \hat{\rho}'(t) &\simeq \frac{i}{\hbar} \int_{-\infty}^t e^{-iH(t-x)/\hbar} [\hat{\rho}_0, U(x)] e^{iH(t-x)/\hbar} dx = \\ &= \frac{i}{\hbar} \int_0^{\infty} e^{-iHt'/\hbar} [\hat{\rho}_0, U(t-t')] e^{iHt'/\hbar} dt', \end{aligned} \quad (37.7)$$

where we set $t' = t - x$. The higher order corrections $\hat{\rho}'', \hat{\rho}''' \dots$ can be obtained in an analogous way.

In the first approximation, formula (37.7) gives an answer to the problem posed. It represents the quantum generalization of the classical formula

(36.6). According to (1.8), the change in the mean value of any quantity described by an operator \hat{B} is of the form

$$\begin{aligned} \langle \Delta \hat{B} \rangle &= \text{Tr}(\hat{\rho}, \hat{B}) - \text{Tr}(\hat{\rho}_0, \hat{B}) = \\ &= \frac{i}{\hbar} \int_0^\infty e^{-iHt'/\hbar} \text{Tr}[\hat{\rho}_0, U(t-t')] e^{iHt'/\hbar} dt'. \end{aligned} \quad (37.8)$$

Formula (37.8) represents a general expression for the response of a system to a small dynamic perturbation. It is approximate in the sense that it takes into account only the first order perturbation. However, in other respects it is applicable to arbitrary systems for any interaction between the particles forming them. As distinct from the kinetic equation, formula (37.8) gives the probability amplitude, i.e. contains the diagonal as well as the non-diagonal matrix elements of the density matrix.

§38. The response of a system to a thermal perturbation

A more complex problem is the response of a system to a thermal perturbation which does not have the character of an external field acting on the system. Since the action of a thermal perturbation cannot be written in the form of an additional term in the Hamiltonian function of an individual particle, the preceding calculations are inapplicable to this case. It turns out, however, that under certain restrictions for thermal perturbations a law of linear response of the type (36.6) can be obtained.

Let us consider a macroscopic system in a non-equilibrium state. We shall characterize the state of this system by a set of macroscopic parameters x_i . We shall determine the value of these parameters assuming their equilibrium values to be the origin.

For a non-equilibrium system the values of the parameters x_i will change in time, so that $\dot{x}_i \neq 0$. We shall restrict ourselves to non-equilibrium systems sufficiently close to equilibrium. This means that the parameters x_i can be considered to be small quantities.

We introduce into consideration the time scale

$$\tau_{\text{micro}} \ll \tau \ll \tau_{\text{macro}}.$$

Here τ_{micro} are times on a microscopic scale characterizing the changes of

state of microscopic parts of the system, and τ_{macro} are times on a macroscopic scale in which a state of total statistical equilibrium is established in the system.

The subsequent discussion will be based on the hypothesis of the existence in the system of the so-called local statistical equilibrium which is established in an intermediate time of the order of τ .

A system in a state of local statistical equilibrium is described by the distribution function ρ^{loc} of the form of the local Gibbs distribution

$$\rho^{\text{loc}} = \frac{1}{Z} \exp \left\{ -\frac{E(p, q) + \Lambda_i x_i}{kT} \right\} d\Gamma. \quad (38.1)$$

The parameters x_i , which are functions of time, change in times of the order of $t \sim \tau_{\text{macro}}$, and for $t \approx \tau_{\text{macro}}$ reduce to zero. Then $\rho^{\text{loc}} \rightarrow \rho^{\text{eq}}$ where ρ^{eq} is the equilibrium Gibbs distribution. Since the x_i 's are small, in the local equilibrium Gibbs distribution they are taken at the value $t = 0$.

The mean values of the parameters x_i can be calculated by the ordinary formulae

$$\bar{x}_i = \frac{1}{Z} \int x_i \exp \left[-(E(p, q) + \Lambda_i x_i)/kT \right] d\Gamma = -kT \frac{\partial \ln Z}{\partial \Lambda_i}. \quad (38.2)$$

The local Gibbs distribution cannot be derived from arbitrary general theoretical propositions. It must be considered as a hypothesis. The validity of this hypothesis is confirmed by numerous experimental facts. The dissatisfaction which arises from such a statement of the problem may be somewhat reduced by recalling the fact that the equilibrium Gibbs distribution is also to a certain degree a hypothesis, even if substantiated and corroborated by numerous plausible considerations.

The parameters Λ_i , conjugate to x_i , also change in times $t \sim \tau_{\text{macro}}$.

The local equilibrium Gibbs distribution allows one to determine the mean energy and entropy of the system (see §21 and §26 of Part III)

$$Z = \exp \left\{ -\frac{E(p, q) + \Lambda_i x_i}{kT} \right\} \Omega(E, x_i), \quad (38.3)$$

$$S(\bar{x}_i) = k \ln \Omega(E, x_i) = \frac{E}{T} + k \ln Z + \frac{\Lambda_i x_i}{T}. \quad (38.4)$$

The entropy is a function of the parameters x_i . In a state of equilibrium,

when $x_i = 0$, i.e. in a time of the order of τ_{macro} the entropy reaches the maximum value

$$S(0) = \frac{E}{kT} + \ln Z. \quad (38.5)$$

Formula (38.4) allows one to express the parameters Λ_i in terms of $S(\bar{x}_i)$. Instead of Λ_i it is convenient to introduce the quantities

$$X_i = \Lambda_i/kT \quad (38.6)$$

so that

$$S(\bar{x}_i) = \frac{E}{kT} + \ln Z + X_i \bar{x}_i. \quad (38.7)$$

We shall assume that the parameters X_i as well as the mean values \bar{x}_i change in times $t \sim \tau_{\text{macro}}$. The values $X_i = 0$ correspond to a state of statistical equilibrium.

The Gibbs distribution, the mean energy, and the entropy of a system change in time. However, as can be seen from definition (38.3), the characteristic time of change of these quantities is of the order of τ_{macro} .

Formula (38.5) allows one to express the parameters X_i in terms of S :

$$X_i = \frac{\partial S}{\partial x_i}.$$

Assuming the local Gibbs distribution to characterize the state of the system at $t = 0$, one can follow the development of the system in time. We shall confine ourselves to times $t \ll \tau_{\text{macro}}$.

The development of a system can be characterized by the quantity

$$I_i(t) = \dot{\bar{x}}_i,$$

representing the mean rate of change of the parameter x_i . Making use of (38.1), we find

$$I_i \approx \int \dot{x}_i(t) \rho^{\text{loc}}(0) d\Gamma = \frac{1}{Z} \int \dot{x}_i(t) \exp [-(E(p, q) + \Lambda_k(0)x_k)/kT] d\Gamma. \quad (38.8)$$

Since the parameters x_i are considered to be small, one can write to a first approximation

$$I_i = \frac{1}{Z} \int \dot{x}_i(t) e^{-E(p,q)/kT} d\Gamma - \sum \Lambda_k(0) \frac{1}{kTZ} \int \dot{x}_i(t) x_k(0) e^{-E(p,q)/kT} d\Gamma =$$

$$= \langle \dot{x}_i(t) \rangle - \frac{1}{kT} \sum \Lambda_k(0) \langle x_k(0) \dot{x}_i(t) \rangle = \langle \dot{x}_i(t) \rangle - \frac{1}{k} \sum X_k(0) \langle x_k(0) \dot{x}_i(t) \rangle.$$

For an equilibrium state, it is obvious that

$$\langle \dot{x}_i(t) \rangle = 0. \quad (38.9)$$

Further, we transform the mean value

$$\langle x_k(0) \dot{x}_i(t) \rangle = \langle x_k(-t) \dot{x}_i(0) \rangle =$$

$$= \langle x_k(0) - \int_{-t}^0 \dot{x}_k(\alpha) d\alpha, \dot{x}_i(0) \rangle d\alpha = \langle x_k(0) \dot{x}_i(0) \rangle - \int_{-t}^0 \langle \dot{x}_i(0) \dot{x}_k(\alpha) \rangle d\alpha.$$

At the initial instant of time the values of the parameters $x_i(0)$ and their rates of change $\dot{x}_i(0)$ are independent of each other. Hence their correlation reduces to zero

$$\langle x_k(0) \dot{x}_i(0) \rangle = 0$$

so that we have

$$\langle \dot{x}_k(0) \dot{x}_i(t) \rangle = - \int_{-t}^0 \langle \dot{x}_i(0) \dot{x}_k(\alpha) \rangle d\alpha = - \int_0^t \langle \dot{x}_i(\alpha) \dot{x}_k(0) \rangle d\alpha. \quad (38.10)$$

Hence

$$I_i(t) = \sum L_{ik} X_k(0) \simeq \sum L_{ik} X_k(t). \quad (38.11)$$

The replacement of $X_k(0)$ by $X_k(t)$ is possible because the thermodynamic forces change in times $\sim \tau_{\text{macro}}$, whereas the time t in formula (38.11) is lim-

ited by the inequality

$$t \ll \tau_{\text{macro}}.$$

L_{ik} denotes the quantities

$$L_{ik} = L_{ki} = \frac{1}{k} \int_0^t \langle \dot{x}_i(0) \dot{x}_k(\alpha) \rangle d\alpha. \quad (38.12)$$

The symmetry of L_{ik} relative to the transposition of indices is obvious. We see that the response to a thermal perturbation is expressed by a linear law. It is the thermodynamic forces X_k that give rise to the thermodynamic fluxes J_i . Thus the coefficients L_{ik} are kinetic coefficients.

The fluxes are characterized by the set of symmetric coefficients L_{ik} . The latter are defined by the correlation functions taken with respect to an equilibrium state of the system. This relation is of a general character, and in this sense the time correlation functions are the basic characteristics of kinetic processes in non-equilibrium systems.

As an example, let us consider the coefficient of diffusion. The mean diffusion path in space

$$\overline{r^2} = \overline{x^2} + \overline{y^2} + \overline{z^2} = 6Dt.$$

Thus by definition the diffusion coefficient

$$D = \lim_{\Delta t \rightarrow 0} \frac{\overline{(\Delta R)^2}}{6\Delta t}. \quad (38.13)$$

We write the displacement $\Delta \mathbf{R}$ in the form

$$\Delta \mathbf{R} = \int_0^{\Delta t} \mathbf{v} dt.$$

Then

$$D = \lim_{\Delta t \rightarrow 0} \frac{1}{6\Delta t} \int_0^{\Delta t} dt'' \int_0^{\Delta t} dt' \overline{\mathbf{v}(t') \mathbf{v}(t'')}. \quad (38.14)$$

The mean value of the velocities does not depend on the zero of time. Hence one can write

$$\overline{v(t') v(t'')} = \overline{v(0) v(t' - t'')} . \quad (38.15)$$

Substituting (38.14) into (38.13) and carrying out one integration, we find directly

$$D = \lim_{\Delta t \rightarrow 0} \frac{1}{3} \int_0^{\Delta t} \left(1 - \frac{t''}{\Delta t}\right) \overline{v(0) v(t'')} dt'' . \quad (38.16)$$

If the interval Δt is microscopically small but nevertheless large in comparison with the characteristic times of molecular processes, then (38.16) can be written in the form

$$D \simeq \lim_{\Delta t \rightarrow 0} \frac{1}{3} \int_0^{\Delta t} dt'' \langle v(0) v(t'') \rangle . \quad (38.17)$$

§39. The calculation of kinetic coefficients. The connection with Boltzmann's equation

The general expression for the response of a system to a dynamic perturbation, found in the preceding section, allows one in principle to find any kinetic coefficients for small perturbations of the system. As an example, we shall confine ourselves to the calculation of the action of an electric field and to the calculation of the electric conductivity*.

We shall consider a quasi-gaseous system made up of a set of charged particles which do not interact with one another (particles of the first kind) and a set of particles of the second kind whose state does not change under the action of an electric field. For example, particles of the second kind may be neutral, or charged but too heavy for their state to be perturbed by a weak field. They may be charged and form a crystal lattice the state of which is also not affected by a weak field, etc.

We have already considered such systems previously by means of Boltzmann's equation (see, for example, §28). However, it should be stressed that

* See M.Lax, Phys. Rev. 109 (1958) 1921.

now we do not assume that the interaction between particles of the first and second kind is weak. On the contrary, the interaction may be arbitrarily strong, and may be described by any law. The Hamiltonian \hat{H} contained in the formulae of the preceding section involves this interaction. The whole set of particles of the first and second kind forms a macroscopic subsystem which as a whole undergoes a weak interaction with the reservoir surrounding it.

We suppose that the subsystem is acted upon by the uniform electric field

$$\mathbf{E} = \lim_{\alpha \rightarrow 0} \mathbf{E}_0 e^{i\omega t} e^{\alpha t}.$$

Then the term U of the total Hamiltonian assumes the form

$$U(t) = - \lim_{\alpha \rightarrow 0} e \mathbf{E}_0 \cdot \mathbf{r} e^{i\omega t} e^{\alpha t}. \quad (39.1)$$

Substituting (39.1) into (37.7), we obtain

$$\hat{\rho}'(t) = -\frac{i}{\hbar} e \mathbf{E}_0 \cdot \int_0^\infty e^{\alpha t'} e^{-i\hat{H}t'/\hbar} [\hat{\rho}_0, \mathbf{r}] e^{i\hat{H}t'/\hbar} e^{-\alpha(t-t')} e^{i\omega(t-t')} dt'.$$

Passing to the limit $\alpha \rightarrow 0$, we have

$$\hat{\rho}'(t) = -\frac{i}{\hbar} e \mathbf{E}_0 \cdot \int_0^\infty e^{-i\hat{H}t'/\hbar} [\hat{\rho}_0, \mathbf{r}] e^{i\hat{H}t'/\hbar} e^{-i\omega t'} dt'. \quad (39.2)$$

Making use of operators in the momentum representation, we get

$$[\hat{\rho}_0, \hat{\mathbf{r}}] = \hat{\rho}_0 \hat{\mathbf{r}} - \hat{\mathbf{r}} \hat{\rho}_0 = i\hbar \left(\hat{\rho}_0 \frac{\partial}{\partial \mathbf{p}} - \frac{\partial}{\partial \mathbf{p}} \hat{\rho}_0 \right) = -i\hbar \frac{\partial \hat{\rho}_0}{\partial \mathbf{p}}, \quad (39.3)$$

Hence

$$\hat{\rho}'(t) = -e \mathbf{E} \cdot e^{i\omega t} \int_0^\infty e^{-i\hat{H}t'/\hbar} \frac{\partial \hat{\rho}_0}{\partial \mathbf{p}} e^{i\hat{H}t'/\hbar} e^{-i\omega t'} dt'. \quad (39.4)$$

The electric current can be written in the form

$$\mathbf{j} = ne \operatorname{Tr} (\hat{\mathbf{v}} \cdot \hat{\rho}'). \quad (39.5)$$

The value of the trace does not depend on the choice of representation. We choose a representation in which the velocity operator is diagonal. Then the expression for the current can be written as follows:

$$\mathbf{j} = ne \int \mathbf{v} \langle \mathbf{v} | \hat{\rho}' | \mathbf{v} \rangle d\mathbf{v}, \quad (39.6)$$

where the matrix $\langle \mathbf{v} | \hat{\rho}' | \mathbf{v} \rangle$ is given by the formula

$$\begin{aligned} \langle \mathbf{v} | \hat{\rho}' | \mathbf{v} \rangle &= -\frac{e}{m} E_0 e^{i\omega t} \cdot \left\langle \mathbf{v} \left| \int_0^\infty e^{-i\omega t'} e^{-i\hat{H}t'/\hbar} \frac{\partial \hat{\rho}_0}{\partial \mathbf{v}} e^{i\hat{H}t'/\hbar} dt' \right| \mathbf{v} \right\rangle = \\ &= -\frac{e E_0 e^{i\omega t}}{m} \cdot \int \int \int_0^\infty e^{-i\omega t'} \langle \mathbf{v} | e^{-i\hat{H}t'/\hbar} | \mathbf{v}' \rangle \left\langle \mathbf{v}' \left| \frac{\partial \hat{\rho}_0}{\partial \mathbf{v}} \right| \mathbf{v}'' \right\rangle \times \\ &\quad \times \langle \mathbf{v}'' | e^{i\hat{H}t'/\hbar} | \mathbf{v} \rangle dt' d\mathbf{v}' d\mathbf{v}'' . \end{aligned} \quad (39.7)$$

Thus the electric conductivity is of the form

$$\begin{aligned} \sigma &= -\frac{ne^2}{m} \int_0^\infty e^{-i\omega t'} dt' \int \mathbf{v} d\mathbf{v} \int d\mathbf{v}' \int d\mathbf{v}'' \times \\ &\quad \times \langle \mathbf{v} | e^{-i\hat{H}t'/\hbar} | \mathbf{v}' \rangle \left\langle \mathbf{v}' \left| \frac{\partial \hat{\rho}_0}{\partial \mathbf{v}} \right| \mathbf{v}'' \right\rangle \langle \mathbf{v}'' | e^{-i\hat{H}t'/\hbar} | \mathbf{v} \rangle . \end{aligned} \quad (39.8)$$

The expression obtained is as general as formula (39.2) for the change of the density matrix. It contains the diagonal ($\mathbf{v}' = \mathbf{v}''$) as well as the non-diagonal ($\mathbf{v}' \neq \mathbf{v}''$) matrix elements of the equilibrium density matrix $\hat{\rho}_0$. Therefore formula (39.8) has a wide range of applicability.

For example, this formula defines the electric conductivity of a liquid metal or a conductor very strongly alloyed with admixtures, when the approximation of Boltzmann's equation is inadequate.

It is very important to compare the theory of response to an external dynamic perturbation based on the exact equation for the density matrix with Boltzmann's kinetic equation. To do this it is easiest to compare the kinetic coefficient we have found, i.e. the electric conductivity (39.8), with an analogous expression obtained by means of Boltzmann's equation. For this we impose a further restriction on the generality of formula (39.8). Namely, we as-

sume that the interaction of charged particles (particles of the first kind) with particles of the second kind is weak. This means that in the first approximation the system of charged particles can be considered to be an ideal gas. Then the density matrix $\hat{\rho}_0$ can be applied to an individual particle, if one sets

$$\hat{\rho}_0 = \frac{e^{-\hat{H}_0/kT}}{Z} = \prod_i \exp(m\hat{v}_i^2/2kT) / \prod_i Z_i, \quad (39.9)$$

where \hat{H}_0 does not involve the interaction of the given particle with all the other particles of the first and second kind.

In the \mathbf{v} -representation, in which expression (39.9) is written, the density matrix $\hat{\rho}_0$ is diagonal. Hence

$$\left\langle \mathbf{v}' \left| \frac{\partial \hat{\rho}_0}{\partial \mathbf{v}} \right| \mathbf{v}'' \right\rangle = \frac{\partial \hat{\rho}_0(\mathbf{v}'')}{\partial \mathbf{v}''} \delta(\mathbf{v}' - \mathbf{v}'') \quad (39.10)$$

and, correspondingly,

$$\begin{aligned} \sigma &= -\frac{ne^2}{m} \int_0^\infty e^{-i\omega t'} dt' \int \mathbf{v} d\mathbf{v} \int d\mathbf{v}' \times \\ &\quad \times \int \langle \mathbf{v} | e^{-i\hat{H}_0 t'/\hbar} | \mathbf{v}' \rangle \left\langle \mathbf{v}' \left| \frac{\partial \hat{\rho}_0}{\partial \mathbf{v}'} \right| \mathbf{v}'' \right\rangle \langle \mathbf{v}'' | e^{i\hat{H}_0 t'/\hbar} | \mathbf{v} \rangle d\mathbf{v}'' = \\ &= -\frac{ne^2}{m} \int_0^\infty e^{-i\omega t} dt' \int \mathbf{v} d\mathbf{v} \int \langle \mathbf{v} | e^{-i\hat{H}_0 t'/\hbar} | \mathbf{v}' \rangle \times \\ &\quad \times \frac{\partial \hat{\rho}_0(\mathbf{v}'')}{\partial \mathbf{v}''} \delta(\mathbf{v}' - \mathbf{v}'') \langle \mathbf{v}'' | e^{i\hat{H}_0 t'/\hbar} | \mathbf{v} \rangle d\mathbf{v}' d\mathbf{v}'' = \\ &= -\frac{ne^2}{m} \int_0^\infty e^{-i\omega t'} dt' \int \mathbf{v} d\mathbf{v} \int W(\mathbf{v}, \mathbf{v}', t') \frac{\partial \hat{\rho}_0(\mathbf{v}')}{\partial \mathbf{v}'} d\mathbf{v}', \end{aligned} \quad (39.11)$$

where

$$W(\mathbf{v}, \mathbf{v}', t') = |\langle \mathbf{v} | e^{-i\hat{H}_0 t'/\hbar} | \mathbf{v}' \rangle|^2. \quad (39.12)$$

It is obvious that $W(\mathbf{v}, \mathbf{v}', t)$ represents the probability that a particle which, with velocity \mathbf{v} at the instant of time $t = 0$, will acquire in time t a velocity \mathbf{v}' . We compare the expression obtained with the general solution of Boltzmann's equation (28.7) in an external force field.

This comparison allows one to convince oneself of their complete identity, provided that ρ_0 is replaced by f_0 . The meaning of the density matrix of a free particle, ρ_0 , is indeed the same as the one-particle distribution function. Thus we see that Boltzmann's equation for a homogeneous quasi-gaseous system indeed follows from the exact equation for the density matrix.

The above reasoning allows a generalization to systems of particles interacting with one another. However, we cannot dwell on this here*.

§40. Onsager's theory

For small departures from an equilibrium state, non-equilibrium processes in a closed system can be described starting from some very general considerations first stated by Onsager.

We shall characterize the state of a closed system by macroscopic parameters x_i . These parameters are functions of time.

For small departures from an equilibrium state, the parameters characterizing the state of the system can be considered to have a thermodynamic meaning. In other words, the x_i should be understood to be the differences between the values of thermodynamic quantities in a given non-equilibrium state and those in an equilibrium state. We recall that in an equilibrium state all thermodynamic quantities have values equal to their means.

It is clear that for large departures from an equilibrium state, thermodynamic concepts make no sense. However, as we have seen above, for small departures from equilibrium, use can be made of thermodynamic quantities, which are then not equal to their mean values. For this it is only necessary that an incomplete local equilibrium exist at each point of the body. For small values of x_i all the quantities characterizing the state of the system and its rate of change can be expanded in series of powers of x_i . In these series one need retain only the first terms, so that one can write

$$\dot{x}_i = \alpha_{ik} x_k, \quad (40.1)$$

* See M. Lax, Phys. Rev. 109 (1958) 1921.

$$S = S_0 - \frac{1}{2} \beta_{ik} x_i x_k, \quad (40.2)$$

$$\dot{S} = -\beta_{ik} \dot{x}_i x_k. \quad (40.2a)$$

Formula (40.1) shows that all processes near an equilibrium state are slow. The entropy of a system in a non-equilibrium state is expressed by the quadratic form (40.2). From the minimum condition it follows that

$$\beta_{ik} = \beta_{ki}. \quad (40.2b)$$

The increase of entropy, S , in unit time is also small. It is obvious that all the formulae given above can be applied to changes of state of the system in limited times t . Namely, on the one hand, these times must be very large in comparison with microscopic times τ_{micro} in order that one may speak of the change of macroscopic quantities.

On the other hand, the system must be in a non-equilibrium state. If total equilibrium is established in it in relaxation time τ_{macro} , then the following inequality must be fulfilled

$$\tau_{\text{micro}} \ll t \ll \tau_{\text{macro}}.$$

We denote by I'_i the microscopic flux

$$I'_i = \dot{x}_i$$

and by X'_i the microscopic thermodynamic force

$$X'_i = \partial S / \partial x_i = -\beta_{ik} x_k.$$

Then the preceding relations can be written in the form

$$I'_i = \alpha_{ik} x_k = -\alpha_{ij} \beta_{jk}^{-1} X'_k = \gamma_{ik} X'_k, \quad (40.3)$$

$$S = S_0 + \frac{1}{2} X'_i x_i, \quad (40.3a)$$

$$\dot{S} = I'_i X'_i. \quad (40.3b)$$

Our further treatment will be based on the following hypothesis of Onsager: macroscopic non-equilibrium state near equilibrium can be considered as a fluctuation. Changes in time of states of a macroscopic equilibrium

system undergoing fluctuations obey the same laws. Let, for example, non-uniform concentration and temperature distributions be produced in a macroscopic system. Then fluxes described by the corresponding macroscopic transport laws will arise in the system. If in an equilibrium system there occur concentration and temperature fluctuations as a result of which the same concentration and temperature distributions are produced, then according to Onsager's hypothesis these fluctuations will resolve according to the same laws as those governing the levelling of concentrations or temperatures in the macroscopic system.

The particle flux and heat flow will be defined by the laws of diffusion and thermal conduction irrespective of how the corresponding concentration and temperature differences arose; either as a result of a spontaneous fluctuation in the equilibrium system, or as a result of external actions which brought the system into the non-equilibrium state.

Thus, according to Onsager's hypothesis, the relation between fluxes and forces, i.e. the macroscopic law

$$I_i = \sum L_{ik} X_k, \quad (40.4)$$

is equally applicable to non-equilibrium systems and to the processes of resolution of fluctuations.

The mean macroscopic fluxes and forces I_i and X_k are obtained by averaging the microscopical fluxes and forces I'_i and X'_k , and the coefficients L_{ik} and γ_{ik} are the same. It turns out that the kinetic coefficients L_{ik} can be expressed in terms of the time correlation function. On the basis of Onsager's hypothesis one can write

$$I'_i(t) = \frac{dx_i}{dt} = L_{ik} X'_k(t). \quad (40.5)$$

We multiply (40.5) by $x_i(0)$, so that

$$x_i(0) \bar{x}_i = x_i(0) L_{ik} X_k(t). \quad (40.6)$$

We now introduce into consideration an ensemble of identical systems differing in the given initial values of the parameters $x_i(0)$. We denote the mean over this ensemble by the symbol $\langle \rangle$.

We find then

$$\langle x_i(0) \bar{x}_i(t) \rangle = L_{ik} \langle x_i(0) X_k(t) \rangle. \quad (40.7)$$

An ensemble of equilibrium closed systems at the initial instant of time forms a Gibbs ensemble. For this ensemble, the probability distribution can be written in the form

$$w(x_1, \dots, x_l \dots) dx_1 \dots dx_N = C \exp [\Delta S(x_1, \dots, x_l \dots)/k] dx_1 \dots dx_N. \quad (40.8)$$

Then the mean values of the quantities involved in formula (40.7) can be written as follows:

$$\langle x_l(0) \overline{\dot{x}_l(t)} \rangle = C \int x_l(0) \frac{d\overline{x_l(t)}}{dt} e^{\Delta S/k} dx_1 \dots dx_l \dots dx_N$$

and, correspondingly,

$$\langle x_l(0) X_k(t) \rangle = C \int x_l(0) X_k(t) e^{\Delta S/k} dx_1 \dots dx_N. \quad (40.9)$$

It should be stressed that, through this, the problem of determining mean values in a non-equilibrium system for which the probability distribution is unknown, turns out, owing to Onsager's hypothesis to be reduced to the problem of calculating the ensemble of means for a Gibbs ensemble of closed systems with a probability distribution given by formula (40.8).

Making use of the quasi-ergodic hypothesis, one can write

$$\left\langle x_l(0) \frac{d\overline{x_l(t)}}{dt} \right\rangle = \langle x_l(0) \dot{x}_l(t) \rangle.$$

Then (40.7) is written in the form

$$\langle x_l(0) \dot{x}_l(t) \rangle = L_{ik} \langle x_l(0) X_k(t) \rangle. \quad (40.10)$$

We transform the left-hand side of (40.10) in the same way as in a preceding section (see (38.9)):

$$\begin{aligned} \langle x_l(0) \dot{x}_l(t) \rangle &= \langle x_l(-t) \dot{x}_l(0) \rangle = \left\langle x_l(0) + \int_0^t \dot{x}_l(\alpha) d\alpha, \dot{x}_l(0) \right\rangle = \\ &= - \int_0^t \langle \dot{x}_l(\alpha) \dot{x}_l(0) \rangle d\alpha. \end{aligned} \quad (40.11)$$

The right-hand side can be transformed in the following way. Since the mean force changes in times $t \sim \tau_{\text{macro}}$, whereas we consider the development of the system in times $t \ll \tau_{\text{macro}}$, we can approximately write

$$X_k(t) \simeq X_k(0).$$

Our problem is to find the mean

$$\langle x_l(0) X_k(0) \rangle = \left\langle x_l(0) \frac{\partial S}{\partial x_k} \right\rangle. \quad (40.12)$$

To calculate the mean we can, on the basis of Onsager's hypothesis, make use of the Gibbs distribution

$$\left\langle x_l(0) \frac{\partial S}{\partial x_k} \right\rangle = C \int x_l(0) \frac{\partial S}{\partial x_k} e^{\Delta S/k} dx_1 \dots dx_N.$$

The integral with respect to x_k can be taken by parts

$$C \int x_l \left(\frac{\partial}{\partial x_k} e^{\Delta S/k} \right) dx_k = -C \int \frac{\partial x_l(0)}{\partial x_k} e^{\Delta S/k} dx_k = -\delta_{lk},$$

since the integrand tends to zero rapidly at the limits. Thus, finally,

$$\left\langle x_l(0) \frac{\partial S}{\partial x_k} \right\rangle = -k\delta_{lk}. \quad (40.13)$$

Substituting (40.11) and (40.13) into (40.7), we find

$$L_{ik} = \frac{1}{k} \left\langle \int_0^t \dot{x}_i(0) \dot{x}_k(\alpha) d\alpha \right\rangle = L_{ki}.$$

We see that Onsager's hypothesis leads to exactly the same expression for the kinetic coefficients as the hypothesis of a local Gibbs distribution (see (38.12)). This proves the equivalence of the two hypotheses.

The symmetry of the kinetic coefficients $L_{ik} = L_{ki}$ has a profound meaning. If, for example, two parameters change and two fluxes arise in the sys-

tem, then from the symmetry property it follows that

$$I_1 = L_{11}X_1 + L_{12}X_2, \quad (40.14)$$

$$I_2 = L_{22}X_2 + L_{12}X_1. \quad (40.15)$$

Formulae (40.14) and (40.15) show that the force X_1 gives a contribution to the flux I_2 , and the force X_2 gives the same contribution to the flux I_1 . Generalization to a large number of forces and fluxes offers no difficulty.

A great number of such cross fluxes is known in physics. As an example, we can point to thermal diffusion and the converse effect of the occurrence of a temperature gradient when gases at the same temperature are mixed. Other examples will be given below.

The symmetry relation allows one to establish the general relation between such cross processes. The use of the symmetry relation makes it possible to describe a great number of connected effects. Agreement of the theory with experimental data is a conclusive proof of Onsager's hypothesis.

§41. Discussion of Onsager's relations

In this and the following sections we shall discuss some consequences of the symmetry of kinetic coefficients. Analysis of these consequences is the basic subject matter of the thermodynamics of irreversible processes. We shall show that from Onsager's relations one can draw conclusions of general character as well as obtaining results of more practical interest. The latter mainly consist of establishing relations between different non-equilibrium processes.

We shall begin with some general consequences. We note, first of all, that Onsager's principle can be obtained from Onsager's hypothesis on the basis of the general theory of fluctuations. In view of the importance of this principle, we also present this more usual derivation.

According to the principle of microscopic reversibility, fluctuations in a closed system are reversible in time, so that for the correlation function one can write

$$\langle x_I(t) x_k(t + \tau) \rangle = \langle x_I(t) x_k(t - \tau) \rangle.$$

On the other hand, changing the zero time on the right-hand side, one can write

$$\langle x_I(t) x_k(t + \tau) \rangle = \langle x_I(t + \tau) x_k(t) \rangle. \quad (41.1)$$

The symbol $\langle \rangle$ in formula (41.1) denotes an average over the ensemble. Averaging once more over time τ , we have

$$\overline{\langle x_l(t) x_k(t + \tau) \rangle}^\tau = \overline{\langle x_l(t + \tau) x_k(t) \rangle}^\tau.$$

Since the two averages are independent and equivalent, subtracting from this equality $\langle x_l(t) \overline{x_k(t)} \rangle$, we get

$$\overline{\langle x_l(t), x_k(t + \tau) - \overline{x_k(t)} \rangle} = \langle x_k(t), \overline{x_l(t + \tau) - x_l(t)} \rangle.$$

Dividing by τ and passing to the limit $\tau \rightarrow 0$, we have

$$\langle x_l(t) \dot{x}_k(t) \rangle = \langle x_k(t) \dot{x}_l(t) \rangle. \quad (41.2)$$

On the basis of Onsager's hypothesis, relation (40.5) holds for fluctuations as well as for macroscopic processes. Its substitution into the last equality gives

$$\langle x_l, L_{ki} X_i \rangle = \langle x_k, L_{li} X_i \rangle. \quad (41.3)$$

But, according to (40.13), we have $\langle x_l X_i \rangle = -k \delta_{li}$, $\langle x_k X_i \rangle = -k \delta_{ki}$ so that (41.3) gives $L_{ki} \delta_{li} = L_{li} \delta_{ki}$. Hence $L_{kl} = L_{lk}$. In this derivation of the proof of the principle of symmetry of kinetic coefficients use was made only of the principle of microscopic reversibility and Onsager's hypothesis (formulae (40.4) and (41.2)). However, here the meaning of the coefficients L_{kl} is not brought to light.

In deriving Onsager's principle we have actually assumed that the system is not placed in a magnetic field and does not rotate. As a matter of fact, if the system is in a magnetic field, under the change of sign of the time $\tau \rightarrow -\tau$ the following equality holds $\mathbf{H} \rightarrow -\mathbf{H}$. For the principle of microscopic reversibility to be fulfilled, the Lorentz force must not change sign. Under time reversal the following equality holds $\dot{x}(\tau) = \dot{x}(-\tau)$. Exactly the same applies to the angular velocity of a rotating body. Reproducing the preceding calculations, one can easily find Onsager's principle to read

$$L_{ik}(\mathbf{H}) = L_{kl}(-\mathbf{H}). \quad (41.4)$$

This equality always holds if the two parameters x_k and x_l are such that under time reversal one of them changes sign whereas the other does not.

Of other general consequences of Onsager's theory we point out the proof of the existence of a dissipative function for mechanical systems performing

in slow motion. One has to add to the equation of motion the components of the force X_k , which for small velocities can be expanded in a series in which only the first term need be retained

$$X_k = \beta_{ik} \dot{q}_i. \quad (41.5)$$

According to (40.26), the tensor β_{ik} is symmetric

$$\beta_{ik} = \beta_{ki}.$$

There is no zero term in the expansion, since the system at rest is not acted upon by dissipative forces. Thus the generalized equation of motion is of the form

$$m\ddot{q}_k = -\frac{\partial U}{\partial q_k} - \beta_{kl} \dot{q}_l.$$

By virtue of the symmetry of β_{kl} , this equality can be written in the form

$$m\ddot{q}_k = -\frac{\partial U}{\partial q_k} - \frac{\partial f}{\partial q_k}, \quad (41.6)$$

where

$$f = \frac{1}{2} \beta_{ik} \dot{q}_i \dot{q}_k. \quad (41.7)$$

The quadratic form f represents the dissipative function. In mechanics, as a rule, the existence of a dissipative function is not proved. However, if the system were acted upon by several frictional forces, then without the principle of symmetry it would be impossible to introduce the dissipative function. Also there is no analogue of the dissipative function for motion in a magnetic field. Although in this case the force is also proportional to the velocity, the tensor β_{kl} is antisymmetric.

§42. Non-equilibrium processes in a one-component system

We shall discuss in detail the theory of non-equilibrium processes in a one-component system in order to elucidate the method of finding thermodynamic forces and fluxes and applications of Onsager's relations. The character of

the results to be obtained will enable us to understand more clearly the merits and shortcomings of the thermodynamics of irreversible processes. Let us consider first of all a one-component closed system consisting of two subsystems with temperatures T_1 and T_2 , pressures p_1 and p_2 , internal energies E_1 and E_2 and numbers of particles N_1 and N_2 per unit volume. Suppose that energy exchange and particle exchange between the subsystems proceed without the appearance of hydrodynamic fluxes. For example, the exchange could be carried out through a porous diaphragm. Let δE_i and δN_i be the changes in the energy and number of particles of one of the subsystems ($i = 1, 2$). Then the change of entropy is

$$\begin{aligned} \delta S_i = & \left(\frac{\partial S}{\partial E} \right)_{N_i} \delta E_i + \left(\frac{\partial S}{\partial N_i} \right)_{E_i} \delta N_i + \\ & + \frac{1}{2} \left[\left(\frac{\partial^2 S}{\partial E_i^2} \right)_{N_i} (\delta E_i)^2 + 2 \frac{\partial^2 S}{\partial E_i \partial N_i} \delta E_i \delta N_i + \left(\frac{\partial^2 S}{\partial N_i^2} \right)_{E_i} (\delta N_i)^2 \right] + \dots \end{aligned}$$

The total change of entropy of the closed system is equal to

$$\begin{aligned} \delta S = \delta S_1 + \delta S_2 = & \frac{1}{2} \left(\frac{\partial^2 S}{\partial E^2} \right) (\delta E)^2 + \frac{1}{2} \left(\frac{\partial^2 S}{\partial N^2} \right) (\delta N)^2 + \left(\frac{\partial^2 S}{\partial E \partial N} \right) \delta N \delta E = \\ = & \frac{1}{2} \left[\frac{\partial}{\partial E} \left(\frac{1}{T} \right) (\delta E)^2 - \frac{\partial}{\partial N} \left(\frac{\mu}{T} \right) (\delta N)^2 + 2 \frac{\partial^2 S}{\partial E \partial N} \delta N \delta E \right]. \end{aligned} \quad (42.1)$$

The linear terms dropped out owing to the laws of conservation of energy and particle number. According to (42.1), the rate of increase of entropy is equal to

$$\begin{aligned} \delta \dot{S} = \delta \dot{E} \left[\frac{\partial}{\partial N} \left(\frac{1}{T} \right) \delta N + \frac{\partial}{\partial E} \left(\frac{1}{T} \right) \delta E \right] - \delta \dot{N} \left[\frac{\partial}{\partial N} \left(\frac{\mu}{T} \right) \delta N + \frac{\partial}{\partial E} \left(\frac{\mu}{T} \right) \delta E \right] = \\ = \delta \dot{E} \delta \left(\frac{1}{T} \right) - \delta \dot{N} \delta \left(\frac{\mu}{T} \right). \end{aligned} \quad (42.2)$$

On the other hand, by virtue of (40.3b) one can write

$$\delta \dot{S} = I_E X_E + I_N X_N, \quad (42.2a)$$

where I_E and I_N are the energy flux and particle flux, and X_E and X_N are the

corresponding thermodynamic forces. Hence we find

$$\begin{aligned} I_E &= \delta \dot{E}, & X_E &= \delta \left(\frac{1}{T} \right) = -\frac{\delta T}{T^2}, \\ I_N &= \delta \dot{N}, & X_N &= -\delta \left(\frac{\mu}{T} \right) = -\frac{v \delta p}{T} + \frac{h \delta T}{T^2}, \end{aligned} \quad (42.3)$$

where v and h are the volume and enthalpy per particle. Taking into account the symmetry of kinetic coefficients, according to (40.4) the fluxes can be written in the form

$$I_E = L_{11}X_E + L_{12}X_N, \quad I_N = L_{12}X_E + L_{22}X_N.$$

Substituting in the value of the forces, we find

$$I_E = \frac{L_{12}h - L_{11}}{T^2} \delta T - \frac{L_{12}v}{T} \delta p, \quad (42.4)$$

$$I_N = \frac{L_{22}h - L_{12}}{T^2} \delta T - \frac{L_{22}v}{T} \delta p. \quad (42.5)$$

Here δT and δp denote the changes in temperature and pressure in passing from one subsystem to the other

$$\delta T = T_1 - T_2; \quad \delta p = p_1 - p_2.$$

It is obvious that the entire treatment makes sense only for small values of δT and δp . The transport of energy and of particles leads to the appearance of cross effects characterized by the kinetic coefficient L_{12} .

An important particular case is that of a process without particle transport. Then formula (42.5) shows that in a stationary state a pressure difference and a temperature difference connected by the relation

$$\delta p = \frac{h - L_{12}L_{22}}{vT} \delta T \quad (42.6)$$

are established between the subsystems. Formula (42.6) defines the so-called thermomolecular pressure. In another important particular case, that in which

the temperature difference $\delta T = 0$, formula (42.4) shows that an energy flux

$$I_E = -\frac{L_{12}v}{T} \delta p = \frac{L_{12}}{L_{22}} I_N = E^* I_N \quad (42.7)$$

flows between the subsystems. This effect, called the thermomechanical effect, shows that a certain energy E^* is transported together with the particles

$$E^* = -L_{12}/L_{22}. \quad (42.8)$$

By means of (42.8) one can write

$$\delta p = \frac{h - E^*}{vT} \delta T. \quad (42.9)$$

Formulae (42.6), (42.7) and (42.9) are in character identical with the thermodynamic relations. They establish very general and often not apparent relations between processes and the measurable quantities characterizing them; in the case given, δp , δT and E^* . However, the values of these quantities must be determined experimentally or calculated by the formulae of kinetic theory. In particular, the value of E^* can easily be calculated for an ideal gas. It is easily seen that for an ideal gas $E^* = h$, so that in this case $\delta p = 0$ for $\delta T \neq 0$.

For systems for which the value of E^* cannot be calculated, it must be determined experimentally. Experiment completely confirms relation (42.9). We note that the determination of forces and fluxes carried out above is not single-valued.

It is clear that expression (42.2a) can be resolved not only in the way it was done in formula (42.3) but also in other ways. It can easily be shown†, however, that this will not affect the final formulae, in particular (42.9).

Let us consider stationary non-equilibrium processes in a certain closed system. In stationary processes the state functions of the system do not depend explicitly on time. It turns out that the entropy production in stationary processes has a minimum value. In order not to complicate our formulae in proving this statement, we shall confine ourselves to the case of two cross

† See S. De Groot, *Thermodynamics of irreversible processes* (North-Holland, Amsterdam, 1952). In presenting the thermodynamics of irreversible processes we follow this monograph.

processes. The entropy production in a body can be written in the form

$$\begin{aligned}
 P &= \int \dot{S} dV = \int I_k X_k dV = \int \sum L_{ik} X_i X_k dV = \\
 &= \int \{L_{11} X_1^2 + L_{12} X_1 X_2 + L_{21} X_1 X_2 + L_{22} X_2^2\} dV = \\
 &= \int \{L_{11} X_1^2 + 2L_{12} X_1 X_2 + L_{22} X_2^2\} dV. \quad (42.10)
 \end{aligned}$$

Substituting the values (42.3) for X_1 and X_2 , we have

$$\begin{aligned}
 P &= \int \{L_{11} [\nabla(1/T)]^2 - 2L_{12} [\nabla(1/T)] [\nabla(\mu/T)] + \\
 &\quad + L_{22} [\nabla(\mu/T)]^2\} dV.
 \end{aligned}$$

The condition for minimum entropy production reads

$$\delta P = \delta \int \dot{S} dV = 0.$$

The variable quantities are the forces $X_1 = \nabla(1/T)$ and $X_2 = \nabla(\mu/T)$. Calculation of the variation gives

$$\begin{aligned}
 [L_{11} \nabla^2(1/T) - L_{12} \nabla^2(\mu/T)] \delta(1/T) &= 0, \\
 [L_{12} \nabla^2(1/T) - L_{22} \nabla^2(\mu/T)] \delta(\mu/T) &= 0, \quad (42.11)
 \end{aligned}$$

or

$$\begin{aligned}
 \nabla \cdot \{L_{11} \nabla(1/T) - L_{12} \nabla(\mu/T)\} &= 0, \\
 \nabla \cdot \{L_{12} \nabla(1/T) - L_{22} \nabla(\mu/T)\} &= 0. \quad (42.12)
 \end{aligned}$$

Formulae (42.11) and (42.12) express the conditions for stationarity. Thus in a stationary state, entropy production has a minimum value.

We stress that the signs in the conditions for symmetry are essential. Hence the proof given is not valid in the presence of a magnetic field.

Among the other general consequences of Onsager's theory we mention the proof of the existence of a dissipative function for mechanical systems undergoing slow motion.

§43. Non-equilibrium processes in many-component systems (diffusion, thermodiffusion, thermoelectric effects)

As another example, let us consider the phenomenon of thermal diffusion. In contrast to the preceding example, we now have to consider a many-component system whose properties depend on spatial coordinates. We shall confine ourselves to the case of chemically non-reacting components, and for simplicity of our formulae we shall assume the mixture to be two-component and isotropic. If, in the system, there are temperature and concentration gradients, then heat and diffusion flows as well as a cross flow, thermal diffusion, will arise in it.

We have already considered thermal diffusion in §24 in the case of a mixture of ideal gases. Comparison of the results of calculations by means of the kinetic theory of gases and the theory of irreversible processes is instructive.

We shall make use of the definitions of forces and fluxes of formulae (40.4) and (42.3).

It should be stressed that, in contrast to the first example, in a system having properties varying continuously in space the motion of particles is always accompanied by the appearance of mass flow. The total mass flow must be equal to zero in a stationary state of the system.

Let I_1 and I_2 be the fluxes of the two components

$$I_1 = L_{11}X_1 + L_{12}X_2 + L_{1E}X_E,$$

$$I_2 = L_{21}X_1 + L_{22}X_2 + L_{2E}X_E,$$

where the forces X_1 , X_2 and X_E are given by formulae identical with (42.3) for each component

$$X_i = \nabla\left(\frac{\mu_i}{T}\right), \quad X_E = \nabla\left(\frac{1}{T}\right).$$

The equality

$$I_1 + I_2 = 0$$

for arbitrary values of forces gives

$$L_{11} = -L_{21}; \quad L_{12} = -L_{22}; \quad L_{1E} = -L_{2E}.$$

Hence

$$\begin{aligned} \mathbf{I}_1 = -\mathbf{I}_2 &= -L_{12} \nabla \left(\frac{\mu_1 - \mu_2}{T} \right) + L_{1E} \nabla \left(\frac{1}{T} \right) = \\ &= -L_{12} \frac{\nabla(\mu_1 - \mu_2)}{T} + \{(L_{1E} - L_{12}(\mu_1 - \mu_2))\} \nabla \left(\frac{1}{T} \right). \end{aligned}$$

From the basic thermodynamic equality we have

$$\begin{aligned} \nabla(\mu_1 - \mu_2) &= \\ &= \frac{\partial(\mu_1 - \mu_2)}{\partial T} \nabla T + \frac{\partial(\mu_1 - \mu_2)}{\partial p} \nabla p + \frac{\partial(\mu_1 - \mu_2)}{\partial c_1} \nabla c_1. \end{aligned}$$

In the presence of mechanical equilibrium $\nabla p = 0$.

Making use of the Gibbs–Duhem relation

$$c_1 d\mu_1 + c_2 d\mu_2 = 0,$$

we finally find

$$\begin{aligned} \nabla(\mu_1 - \mu_2) &= -(s_1 - s_2) \nabla T + \frac{1}{c_2} \left(\frac{\partial \mu_1}{\partial c_1} \right) \nabla c_1 = \\ &= (s_1 - s_2) T^2 \nabla \left(\frac{1}{T} \right) + \frac{1}{1 - c_1} \left(\frac{\partial \mu_1}{\partial c_1} \right) \nabla c_1. \end{aligned}$$

Then for the flux \mathbf{I}_1 we obtain

$$\mathbf{I}_1 = -\frac{L_{12}}{T(1 - c_1)} \left(\frac{\partial \mu_1}{\partial c_1} \right) \nabla c_1 - \frac{L_{1E} - L_{12}(h_1 - h_2)}{T^2} \nabla T. \quad (43.1)$$

If we introduce the coefficient of diffusion

$$D_{12} = L_{12} \frac{1}{(1 - c_1) T} \frac{\partial \mu_1}{\partial c_1} \quad (43.2)$$

and the coefficient of thermal diffusion

$$D_T = \frac{L_{1E} - L_{12}(h_1 - h_2)}{T^2(1 - c_1)c_1}, \quad (43.3)$$

then (43.1) gives

$$\mathbf{I}_1 = -D_{12}\nabla c_1 - D_T c_1(1 - c_1)\nabla T. \quad (43.4)$$

The sign of the coefficients of diffusion and thermal diffusion is defined by the requirement that the rise of entropy be positive. The relations obtained can be applied to any binary mixtures in the isotropic (liquid or gaseous) phase, for any values of the concentration, but for small values of temperature and concentration gradients. Similarly one can find expressions for the flux in the presence of external forces (for example, when there is ion migration in an electric field, viscosity, or chemical reactions between the components).

Generalization to the case of many-component systems offers no difficulty.

We stress that in our derivation we have assumed that there is no thermomechanical effect in a system with continuously varying parameters (i.e. that $\nabla p = 0$). This constancy of pressure always exists when there is no mean flux of matter $\mathbf{I}_1 + \mathbf{I}_2 = 0$ and the system is described by macroscopic laws. Comparing the result obtained with the conclusions of §24, we see clearly that their interrelation is the same as that between the results of thermodynamic and statistical descriptions. From the kinetic theory of gases the molecular meaning of all the quantities was found, but only for the simplest case of an ideal gas of small concentration ($c_1 \ll 1$). In the thermodynamics of irreversible processes one obtains very general relations, but they contain constants whose meaning and values remain unknown and which must be determined from experimental data.

As the last and perhaps the least trivial example of the application of the thermodynamics of irreversible processes, we cite the theory of thermoelectric phenomena. Let us consider phenomena arising in a thermocouple made of two different metallic conductors whose junctions are maintained at different temperatures T_1 and T_2 . E.m.f. is applied to the circuit, so that there are heat flux and electric current in the system. We shall assume that the processes are of a stationary character. We can write for the fluxes the general

expressions

$$I_e = L_{11} \delta \left(\frac{\mu}{T} \right) + L_{12} \delta \left(\frac{1}{T} \right), \quad (43.5)$$

$$I_E = L_{21} \delta \left(\frac{1}{T} \right) + L_{22} \delta \left(\frac{\mu}{T} \right). \quad (43.6)$$

In an electric field (see §59 of Part III)

$$\mu = \mu_0 + e\varphi,$$

where φ is the field potential. Therefore (43.5) and (43.6) are conveniently written in the form

$$I_e = -L'_{11} \frac{\delta\varphi}{T} - L'_{12} \frac{\delta T}{T^2}, \quad (43.7)$$

$$I_E = -L'_{21} \frac{\delta\varphi}{T} - L'_{22} \frac{\delta T}{T^2}, \quad (43.8)$$

where $\delta\varphi$ is the e.m.f., and $\delta T = T_1 - T_2$.

Onsager's relation reads

$$L'_{12} = L'_{21}.$$

The coefficients in eqs. (43.7) and (43.8) can be expressed in terms of the electrical conductivity σ and the thermal conductivity λ . Namely, setting $\delta T = 0$ we find

$$I_e = -\frac{L'_{11}}{T} \delta\varphi = \sigma \delta\varphi, \quad (43.9)$$

so that $\sigma = -L'_{11}/T$. Setting $I_e = 0$, $\delta T \neq 0$, we have

$$0 = \sigma \delta\varphi - L'_{12} \frac{\delta T}{T^2}, \quad (43.10)$$

$$I_E = -L'_{12} \frac{\delta\varphi}{T} - L'_{22} \frac{\delta T}{T^2} = \lambda \delta T. \quad (43.11)$$

Hence

$$\lambda = \{(L'_{12})^2 - L'_{11}L'_{22}\}/L'_{11}T^2.$$

Consider the particular cases of formulae (43.9) and (43.11). Let the thermojunction be heated in an open circuit ($I_e = 0$). Then a thermoelectromotive force

$$\delta\varphi = -\frac{L'_{12}}{L'_{11}T}\delta T \quad (43.12)$$

arises in the circuit. The appearance of a potential difference in the open circuit is called the Seebeck effect. When current flows between different conductors under isothermal conditions ($\delta T = 0$), energy transport occurs and a certain amount of heat, called the Peltier heat, is released. Setting $\delta T = 0$ in (43.9) and (43.11) we obtain

$$I_E = \frac{L'_{21}}{L'_{11}}I_e = \Pi_{12}I_e, \quad (43.13)$$

where Π_{12} is the heat released when the current $I_e = 1$ flows.

Comparing (43.13) and (43.12) and making use of Onsager's reciprocity relation, we find

$$\frac{\delta\varphi}{\delta T} = -\frac{\Pi_{12}}{T}. \quad (43.14)$$

This formula is called the second Thomson relation. It contains only quantities which may be directly measured and is in good agreement with experiment. It should be noted that for its derivation it is necessary to use the reciprocity relation. In the book by de Groot mentioned above it is shown very clearly how the implicit use of this relation allowed Thomson to obtain formula (43.14) from thermodynamic considerations which are inapplicable explicitly to the phenomena in question.

In conclusion, we note that the fact that thermal conductivity is an essentially positive quantity requires that the inequality $L'_{11}L'_{22} - (L'_{12})^2$ be fulfilled. The proof of this inequality is given in the book by de Groot.

§44. The fluctuation-dissipation theorem

In §29 of Part IV we have already discussed the fluctuation-dissipation theorem and pointed out that the region of applicability of this theorem is much greater than follows from the proof given there. The quantum-mechanical derivation of the fluctuation theorem which we are going to develop shows that the theorem is applicable for frequencies and temperatures where quantum effects assume a fundamental importance.

There are many different ways in which the fluctuation-dissipation theorem can be proved. We have chosen the simplest and most direct of them*.

Let a system of particles in a reservoir be acted upon by a dynamic perturbation (an external force) varying in time according to a harmonic law.

For concreteness we assume that the perturbing force is

$$\mathbf{F} = e\mathbf{E}_0 e^{i\omega t}, \quad (44.1)$$

where \mathbf{E}_0 is an uniform electric field. To the perturbation (44.1) there corresponds a term in the Hamiltonian of the form

$$\hat{H}' = e\mathbf{E}_0 \cdot \mathbf{r} e^{i\omega t}. \quad (44.2)$$

It should be stressed that the subsequent proof is valid for any pair of quantities in the Hamiltonian for which the perturbation is of the form

$$\hat{H}' = (\hat{\xi}\mathbf{F}). \quad (44.3)$$

Here $\hat{\xi}$ is the operator corresponding to a certain parameter ξ . As to the assumptions that the field is uniform and varies according to a harmonic law, one can always expand (44.3) in a Fourier integral and consider the action of each harmonic. Hence we do not actually reduce the generality when assuming a perturbation of the form (44.3).

The time-dependent perturbation \hat{H}' gives rise to transitions in the system. As a result of this, energy is absorbed in the system, being dissipated into heat.

The energy absorbed per unit time, i.e. the absorbed power can be written in the form

$$Q = \left\langle \left(\sum_{n>m} \hbar\omega_{nm} w_{mn} + \sum_{n<m} \hbar\omega_{mn} w_{nm} \right) \right\rangle. \quad (44.4)$$

* L.D. Landau and E.M. Lifshitz, *Statistical physics* (Pergamon Press, Oxford, 1958).

The first term represents the absorbed power, and the second, the emitted power. The averaging is carried out over the equilibrium Gibbs distribution. Since the probabilities of direct transitions are equal to those of inverse transitions and are given by formula (56.8) of Part V, we obtain

$$Q = \left\langle \frac{\pi}{2\hbar^2} \sum |H'_{nm}|^2 [\delta(\omega - \omega_{nm}) \hbar \omega_{nm} + \delta(\omega + \omega_{nm}) \hbar \omega_{nm}] \right\rangle = \\ = \frac{\pi e^2 E_0^2}{2\hbar} \langle |r_{nm}|^2 \omega_{nm} [\delta(\omega - \omega_{nm}) + \delta(\omega + \omega_{nm})] \rangle.$$

The Gibbs statistical averaging gives

$$Q = \frac{\pi e^2 E_0^2}{2\hbar Z} \left\{ \sum_n \sum_m \exp(-\epsilon_n/kT) |r_{nm}|^2 \omega_{nm} \delta(\omega - \omega_{nm}) + \right. \\ \left. + \sum_n \sum_m \exp(-\epsilon_n/kT) |r_{nm}|^2 \omega_{nm} \delta(\omega - \omega_{mn}) \right\}.$$

Interchanging the summation indices in the second term, we get

$$Q = \frac{\pi e^2 E_0^2}{2\hbar Z} \left\{ \sum_n \sum_m |r_{nm}|^2 \omega_{nm} \delta(\omega - \omega_{nm}) \times \right. \\ \times \{ \exp(-\epsilon_n/kT) - \exp(-\epsilon_m/kT) \} = \\ = \frac{\pi e^2 E_0^2}{2\hbar Z} (1 - e^{-\hbar\omega/kT}) \omega \sum_n \sum_m \exp(-\epsilon_n/kT) |r_{nm}|^2 \delta(\omega - \omega_{nm}) \Big\}. \quad (44.5)$$

Instead of the dissipated power Q , one often finds the imaginary part of the dielectric susceptibility, related to the dissipated power by formula (31.38) of Part IV

$$Q = \frac{1}{8\pi} \omega \epsilon^{\text{im}} E_0^2. \quad (44.6)$$

From (44.5) and (44.6) we obtain

$$\epsilon^{\text{im}}(\omega) = \frac{\pi}{\hbar Z} (1 - e^{-\hbar\omega/kT}) \sum \sum \exp(-\epsilon_n/kT) \times \\ \times |r_{nm}|^2 \delta(\omega - \omega_{nm}). \quad (44.7)$$

We have supplied the imaginary part ϵ^{im} with the argument ω in order to emphasize its frequency dependence. It turns out that the quantities $\epsilon^{\text{im}}(\omega)$ or Q can be expressed in terms of the time correlation function of the operator \hat{r} , more precisely in terms of the Fourier component of this function. The classical definition of the correlation function must be somewhat modified, since the values of the operator in the Heisenberg representation, taken at different instants of time, are in general not commutative and $\hat{r}(t) \hat{r}(t + \tau) \neq \hat{r}(t + \tau) \hat{r}(t)$. On the other hand, the times t and $t + \tau$ are completely equivalent. Therefore it is natural to define the correlation function of the time-dependent quantum-mechanical operator \hat{r} as the symmetrized product $\hat{r}(t) \hat{r}(t + \tau)$ averaged after Gibbs, i.e.,

$$\begin{aligned} \frac{1}{2} \langle \hat{r}(t) \hat{r}(t + \tau) + \hat{r}(t + \tau) \hat{r}(t) \rangle &= \\ &= \frac{1}{2} \text{Tr} \{ \hat{\rho}_0, \hat{r}(t) \hat{r}(t + \tau) + \hat{r}(t + \tau) \hat{r}(t) \} = \\ &= \frac{1}{2Z} \text{Tr} \{ \exp(-H/kT), (\hat{r}(t) \hat{r}(t + \tau) + \hat{r}(t + \tau) \hat{r}(t)) \}. \end{aligned}$$

The time dependence of the operator is given by formula (49.8) of Part V. Hence

$$\begin{aligned} (\hat{r}(t) \hat{r}(t + \tau))_{nm} &= e^{iH_0 t/\hbar} \hat{r} e^{iH_0 t/\hbar} e^{-iH_0(t+\tau)/\hbar} \hat{r} e^{iH_0(t+\tau)/\hbar} = \\ &= \sum \exp(i\omega_{nm} \tau) |r_{nm}|^2. \end{aligned}$$

Thus one can write

$$\begin{aligned} \frac{1}{2} \langle \hat{r}(t) \hat{r}(t + \tau) + \hat{r}(t + \tau) \hat{r}(t) \rangle &= \\ &= \frac{1}{2Z} \sum \exp(-\epsilon_n/kT) |r_{nm}|^2 (\exp(i\omega_{nm} \tau) + \exp(i\omega_{mn} \tau)). \quad (44.8) \end{aligned}$$

In order to relate the correlation function directly to the dissipated power, one has to eliminate factors of the type $\exp(i\omega_{nm} \tau)$ in the last sum.

For this we pass from the correlation function to its Fourier component.

According to the Wiener-Khinchin theorem (see §5 of Part III), we have

$$\begin{aligned}
 g(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{2} \langle \tilde{\chi}(t) \tilde{\chi}(t+\tau) + \tilde{\chi}(t+\tau) \tilde{\chi}(t) \rangle e^{-i\omega\tau} d\tau = \\
 &= \frac{1}{4\pi Z} \sum \exp(-\epsilon_n/kT) |r_{nm}|^2 \times \\
 &\times \int_{-\infty}^{\infty} \exp[i(\omega_{nm} - \omega)\tau] + \exp[i(\omega_{mn} - \omega)\tau] d\tau = \\
 &= \frac{1}{2Z} \sum_{n,m} \exp(-\epsilon_n/kT) |r_{nm}|^2 [\delta(\omega - \omega_{nm}) + \delta(\omega - \omega_{mn})] .
 \end{aligned}$$

Interchanging the summation indices in the second term, we finally obtain

$$\begin{aligned}
 g(\omega) &= \frac{1}{2Z} \sum_m \sum_n \exp(-\epsilon_n/kT) (1 + \exp[-(\epsilon_m - \epsilon_n)/kT]) \times \\
 &\times |r_{nm}|^2 \delta(\omega - \omega_{nm}) = \frac{1 + e^{-\hbar\omega/kT}}{2Z} \times \\
 &\times \sum \sum \exp(-\epsilon_n/kT) |r_{nm}|^2 \delta(\omega - \omega_{nm}) . \quad (44.9)
 \end{aligned}$$

Comparing (44.7) and (44.9) we see that the following direct relation exists between $\epsilon^{\text{im}}(\omega)$ and $g(\omega)$:

$$\begin{aligned}
 \epsilon^{\text{im}}(\omega) &= \frac{\hbar}{2\pi} \frac{1 - e^{-\hbar\omega/kT}}{1 + e^{\hbar\omega/kT}} g(\omega) = \\
 &= \frac{\hbar}{2\pi} \cotan\left(\frac{\hbar\omega}{2kT}\right) g(\omega) . \quad (44.10)
 \end{aligned}$$

Formula (44.10) represents the quantum generalization of Nyquist's formula. In the classical limit $\hbar\omega \ll kT$ it goes over into the classical Nyquist formula.

We recall that in the case of thermodynamic fluctuations the fluctuation-dissipation theorem allows one to relate the spectral density fluctuation with the directly measured quantity $\epsilon^{\text{im}}(\omega)$.

Solid-state Theory

§45. A solid body as a quantum-mechanical system

Solid-state theory turns out to be one of those fields of physics where the application of quantum-mechanical concepts has proved to be particularly fruitful. Only on this basis has it been possible to formulate the principles of solid-state theory, to create the theory of the kinetic and equilibrium properties of metals, semiconductors and dielectrics, to understand the essential differences between them, and to explain numerous and peculiar phenomena in solid bodies which for a long time seemed paradoxical (for example, superconductivity or ferromagnetism).

At the present time solid-state quantum theory has reached a stage of development such that it allows one to predict new, detailed and original phenomena in solid bodies.

At first sight the fact that quantum effects can manifest themselves in solid bodies, which are macroscopic objects, appears to be paradoxical. However, it should be recalled that every monocrystal represents in essence a gigantic molecule (see §49 of Part III). Hence the thermal, electrical and other properties of solid bodies are underlain by quantum effects. We saw this in chapter 7 of Part III, where some qualitative foundations of crystal-lattice theory were presented. In this book we shall naturally confine ourselves to the presentation of the general principles of solid-state quantum theory.

For a more detailed acquaintance with the subject, the reader is referred to more specialized textbooks*.

Any macroscopic solid body represents a system of an enormous number of particles interacting strongly with each other. It is therefore clear that for the construction of solid-state theory it is necessary not only to combine quantum-mechanical and statistical descriptions, but also essentially to simplify and schematize the picture of the interactions existing between the particles of solid-body. This means that it is necessary to create a solid-body model which is sufficiently adequate for the object, yields its basic features but does not take into account minor and inessential details.

Our further considerations will be based on the following model: a solid body represents a set of ions and valence electrons. By ions we understand atomic nuclei together with all the electrons in closed shells. The interaction of the electrons of the closed shells with the nucleus is so strong that the proximity of other atoms and the crystal form have no significant effect on it.

We shall abstract the internal structure of ions and assume that the electrons of the closed shells of each ion interact only with their own nucleus. We shall consider ions to be point particles possessing masses M (identical in the case of elementary substances and different in the case of chemical compounds).

It is clear, however, that when ions draw together to distances of the order of their own size, the valence electrons of a given atom enter into a strong interaction with neighbouring nuclei and their electron shells. This interaction ensures the chemical binding between ions, which in the case of crystals is called the cohesive force. Hence valence electrons cannot be considered to be localized in a given atom, and in certain cases they may move throughout the crystal. The presence of valence electrons is not a necessity for a solid body. For example, they are absent in crystals of elements of the zeroth group of the periodic table. In such crystals the binding between the atoms forming the lattice arises from the van der Waals force. In covalent crystals, there are not only ions at lattice points but also neutral atoms. Therefore, instead of speaking of ions, one frequently speaks of nuclei at crystal lattice points. However, in the overwhelming majority of phenomena occurring in solid bodies, electrons play a fundamental role. Therefore we shall consider the most general case where the crystal contains ions and valence electrons.

* We particularly recommend: J.M.Ziman, *Principles of the theory of solids* (Cambridge University Press, Cambridge, 1964); R.Peierls, *Quantum theory of solids* (Clarendon Press, Oxford, 1955); C.Kittel, *Quantum theory of solids* (Wiley, New York, 1963).

Let \mathbf{R}_i denote the coordinates of the ions (nuclei), and \mathbf{r}_k the coordinates of the valence electrons. Then the Hamiltonian of the system of ions and electrons can be written in the form

$$\begin{aligned} \hat{H} = & - \sum_i \frac{\hbar^2}{2M} \nabla_i^2 - \sum_k \frac{\hbar^2}{2m} \nabla_k^2 + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \\ & + \sum_{i > j} U(\mathbf{R}_i - \mathbf{R}_j) + \sum_{i > j} U(\mathbf{r}_i - \mathbf{R}_j). \end{aligned} \quad (45.1)$$

The first sum (over all the ions) represents the kinetic energy of the ions. The second sum gives the kinetic energy of all the valence electrons. The three last terms describe respectively the Coulomb interaction between the electrons, the interaction between the ions, and the interaction between the ions and the electrons. These interactions depend on the distance between the corresponding particles. In the first two interaction sums the particle indices must be different, while in the third interaction sum they may be the same and the factor $\frac{1}{2}$ is absent.

Solid-state theory is based on the assumption that it is possible to use adiabatic perturbation theory. The set of valence electrons is considered to be a fast subsystem, and the set of nuclei and associated closed shell electrons a slow subsystem. In this approximation the total wave function of the system, $\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots; \mathbf{R}_1, \mathbf{R}_2, \dots)$, can be written in the form

$$\begin{aligned} \Psi(\mathbf{r}_1, \mathbf{r}_2, \dots; \mathbf{R}_1, \mathbf{R}_2, \dots) = \\ = \Psi_I(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) \Psi_{\text{el}}(\mathbf{r}_1, \mathbf{r}_2, \dots; \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N), \end{aligned} \quad (45.2)$$

where Ψ_I is the wave function of the system of ions, and Ψ_{el} is the wave function of the system of electrons. In the adiabatic approximation these wave functions satisfy the equations (see (57.7) and (57.8) of Part V)

$$\begin{aligned} \left\{ -\frac{\hbar^2}{2m} \sum_k \nabla_k^2 + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i > j} U(\mathbf{r}_i - \mathbf{R}_j) \right\} \Psi_{\text{el}} = \\ = E_{\text{el}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) \Psi_{\text{el}}, \end{aligned} \quad (45.3)$$

$$\left\{ -\frac{\hbar^2}{2M} \sum_i \nabla_i^2 + \sum_{i > j} U(\mathbf{R}_i - \mathbf{R}_j) + E_{\text{el}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) \right\} \Psi_I = E \Psi_I. \quad (45.4)$$

In this approximation the equation for the wave function of the fast subsystem does not involve the momenta (derivatives with respect to coordinates) of particles of the slow subsystem. This means that the electrons move relative to given positions of the heavy ions. To any change of the latter there corresponds a redistribution of electrons which always manage to readjust themselves to the changed arrangement of ions. Hence the potential energy of interaction between the electrons and ions involves the coordinates of the ions only as parameters. The energy of the valence electrons is in its turn a constituent part of the potential energy of the system of ions.

The application of the adiabatic approximation to the treatment of solids has been recently substantiated and the limits of its applicability determined.

In the adiabatic approximation, one succeeds to a certain degree in separating the motion of the electrons and ions. Knowing the energy of interaction of the electrons and ions and considering the positions of the ions to be given, one can in principle solve eq. (43.3) and find the energy E_{el} of the system of electrons which depends, of course, on the positions of the ions. This value of E_{el} must then be used in solving eq. (45.4).

It is clear, however, that the actual carrying out of this program is associated with insuperable difficulties, since each of eqs. (45.3) and (45.4) contains a macroscopically large number of variables. Therefore modern solid-state theory is associated with the consideration of simplified models, reflecting the most important features of their behaviour.

§46. The crystal lattice

We begin with the discussion of the properties of the slow subsystem, i.e. with the properties of the system of ions. Eq. (45.4) completely defines the behaviour of this system. Its solution must, first of all, give an answer to the question as to how a solid body, i.e. a body with a regular arrangement of heavy particles at crystal lattice points, is formed. It is clear, however, that the exact solution of this equation is out of question. Hence the problem of calculating the cohesive forces keeping the crystal together, and of elucidating the role played in them by van der Waals and exchange forces as well as by the valence electrons existing in the system, is the most difficult of solid-state theory. It cannot as yet be considered solved, although very important relevant results have been obtained.

We cannot discuss this problem here, and shall confine ourselves only to qualitative considerations. It is clear, first of all, that the interaction between ions (or atoms in the case of covalent crystals) ensures, in principle, the exis-

tence of a crystal structure. To the minimum energy of the system there corresponds their regular arrangement in space at certain distances from one another. In the case of covalent crystals an exchange interaction, which gives rise to attraction at large distances and very strong repulsion at small distances, takes place between atoms. In the case of ionic crystals, the attraction is of an electrostatic character, whereas the repulsion is due to a complex interaction between atomic residues which arise at small distances. In metals, a substantial contribution to the cohesive forces is given by valence electrons which considerably weaken the repulsion between ions. As a result of the interaction, atoms or ions are distributed at distances corresponding to the minimum energy. These distances are similar to those between atoms in molecules ($1-2\text{\AA}$). Correspondingly, the cohesive energy per atom is of the same order of magnitude as the energy of chemical binding, and varies from 1 eV (for metals) up to 10 eV (for ionic crystals of the sodium chloride type) per particle.

The regular arrangement of ions or atoms at crystal lattice points plays a very important role and determines many physical properties of solids. Therefore in what follows we shall need the elements of crystal lattice geometry. The basic geometric property of a crystal lattice is its translational symmetry. We shall neglect the finite size of the crystal and shall assume that it fills all space. The property of translational symmetry allows one to introduce the three basis vectors \mathbf{a} , \mathbf{b} , \mathbf{c} . When the crystal translates in space by the vector

$$\mathbf{n} = n_1 \mathbf{a} + n_2 \mathbf{b} + n_3 \mathbf{c}, \quad (46.1)$$

where n_1 , n_2 and n_3 are integers, it coincides with itself (here it is assumed that the origin of the vector \mathbf{n} coincides with one of the lattice points).

The basis vectors define the main crystallographic directions in which translations can be carried out. This can be formulated also in a somewhat different way. From the three basis vectors \mathbf{a} , \mathbf{b} , \mathbf{c} we construct a parallelepiped. This parallelepiped is called a unit cell. Then the entire crystal of infinite size is obtained by infinite reproduction of unit cells. It should be noted that the choice of basis vectors is not itself quite unambiguous. This is most easily seen in fig. VI.6 in which a simple cubic lattice is shown. One can choose as basis vectors the three vectors \mathbf{a} , \mathbf{b} , \mathbf{c} or the three vectors \mathbf{a}' , \mathbf{b}' , \mathbf{c}' .

We see that if the atoms of the lattice are of one type, then all the lattice points are obtained by the translation of one lattice point by the vector \mathbf{n} for different values of the integers n_1 , n_2 and n_3 . In this case the unit cell is said to contain one atom (at a lattice point). Such a lattice is called a Bravais lattice.

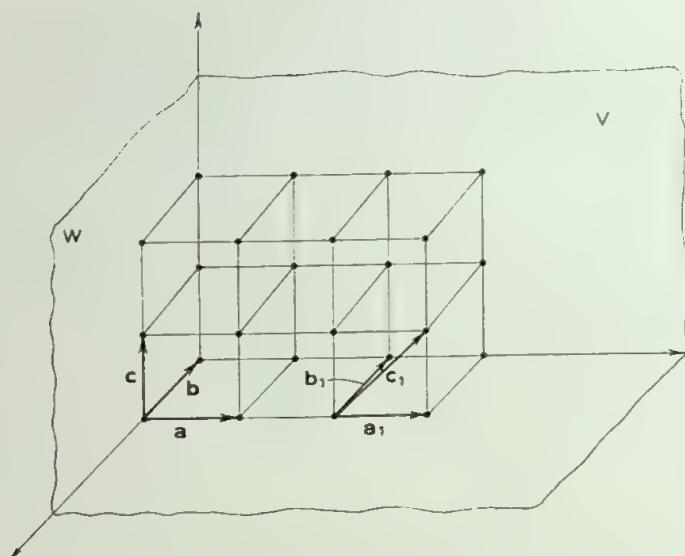


Fig. VI.6

If the unit cell contains two or more atoms, then the crystal is said to have a lattice with a basis. A lattice with a basis represents a set of simple lattices pushed against each other. For example, a unit cell of the sodium chloride crystal contains a sodium atom and a chlorine atom. The sodium and chlorine atoms form simple cubic lattices shifted with respect to each other by one half of the cube edge. The magnitudes of the basis vectors are called lattice constants. The lattice constant represents the distance (in the given main crystallographic direction) between nearest neighbours.

We shall not dwell on the discussion of the possible symmetry elements and crystal classes. We only point out that translation transformations are not the only symmetry transformations which bring the crystal into correspondence with itself. Other such symmetry transformations are rotations and reflections.

In what follows we shall frequently encounter the concept of the reciprocal lattice. The reciprocal lattice is a lattice constructed on the basis vectors a_1 , b_1 and c_1 defined by the formulae

$$a_1 = 2\pi \frac{(b \times c)}{a \cdot (b \times c)}; \quad b_1 = 2\pi \frac{(c \times a)}{a \cdot (b \times c)}; \quad c_1 = 2\pi \frac{(a \times b)}{a \cdot (b \times c)}. \quad (46.2)$$

The reciprocal lattice vector \mathbf{K} is connected with the basis vectors \mathbf{a}_1 , \mathbf{b}_1 and \mathbf{c}_1 by the relation

$$\mathbf{K} = \mathbf{a}_1 m_1 + \mathbf{b}_1 m_2 + \mathbf{c}_1 m_3, \quad (46.3)$$

where m_1 , m_2 and m_3 are integers.

It is obvious that

$$\mathbf{K} \cdot \mathbf{a} = 2\pi m_1, \quad \mathbf{K} \cdot \mathbf{b} = 2\pi m_2, \quad \mathbf{K} \cdot \mathbf{c} = 2\pi m_3. \quad (46.4)$$

The importance of the reciprocal lattice vector is associated with the fact that the following equality always holds

$$\mathbf{K} \cdot \mathbf{n} = 2\pi(n_1 m_1 + n_2 m_2 + n_3 m_3) = 2\pi N, \quad (46.5)$$

where N is an integer.

Because of this, the frequently encountered expression $e^{i\mathbf{K} \cdot \mathbf{r}}$ reduces to unity for $\mathbf{r} = \mathbf{n}$, i.e. when the vector \mathbf{r} coincides with one of the lattice points,

$$e^{i\mathbf{K} \cdot \mathbf{n}} = e^{i2\pi N} = 1. \quad (46.6)$$

§47. Lattice vibrations

Ions or atoms located at crystal lattice points are in thermal motion and vibrate about equilibrium positions. In solid crystals at temperatures below the melting point the amplitude of these vibrations is small in comparison with the lattice constant.

We denote the position of the lattice point (equilibrium position) of the n th ion by \mathbf{n} , and the displacement of the n th atom by ξ_n , so that

$$\mathbf{R}_n = \mathbf{n} + \xi_n. \quad (47.1)$$

Then proceeding in exactly the same way as we did in §50 of Part III in the classical approximation, we can expand the potential energy of interaction $U(\mathbf{R}_i - \mathbf{R}_j)$ in a series in powers of small displacements. We saw in §50 of Part III that the character of the motion of a lattice depends essentially on its structure. The presence alone of a cell with a basis (for example, the presence of two kinds of particles having different masses) gives rise to new vibrational modes.

Since we are in the first place interested in the fundamental theoretical aspects of the subject, we shall not consider complicated calculations and shall restrict ourselves to crystals with a simple Bravais lattice. Moreover, for simplicity of notation we shall drop the vector indices, as if the crystal were a one-dimensional chain. Then near the equilibrium positions one can write that

$$\begin{aligned}
 U(R_n - R_{n'}) &= U(n - n') + \frac{1}{2} \sum_{n', n} \frac{\partial^2 U}{\partial R_n \partial R_{n'}} \xi_n \xi_{n'} + \\
 &+ \frac{1}{6} \sum_{n, n', n''} \frac{\partial^2 U}{\partial R_n \partial R_{n'} \partial R_{n''}} \xi_n \xi_{n'} \xi_{n''} + \dots
 \end{aligned} \quad (47.2)$$

Retaining only the first two terms of the expansion, the Hamiltonian can be written in the form

$$H = H_0 + \sum_n P_n^2 / 2M + \sum_{n, n'} a_{nn'} \xi_n \xi_{n'}, \quad (47.3)$$

where H_0 involves all terms which do not depend on displacements.

The equation of motion reads

$$\dot{P}_n = M \ddot{\xi}_n = - \sum_{n'} a_{nn'} \xi_{n'}. \quad (47.4)$$

The solutions of the equation of motion which satisfy the translational symmetry conditions are the functions

$$\xi_n = \sum_f q_f e^{i(fn - \omega f t)}. \quad (47.5)$$

Substituting (47.5) into (47.4) gives

$$M \omega_f^2 = \sum_{n'} a_{nn'} e^{if(n' - n)}. \quad (47.6)$$

We recall that the values of the wave number are determined from boundary conditions. If conditions of periodicity over the length $L = Na$ are chosen to

be the boundary conditions, then

$$f = \pm \frac{\pi m}{aN} \quad (m = 1, 2, \dots, N). \quad (47.7)$$

The wave number f in the lattice is not determined uniquely but to within the quantity $2\pi/a$; that is the displacement does not change under the replacement $f \rightarrow f + 2\pi/a$. We shall normalize the wave numbers in the interval

$$-\pi/a \leq f \leq \pi/a. \quad (47.8)$$

In a one-dimensional chain the reciprocal lattice vector is equal to $2\pi/a$. Without reproducing the calculations of § 50 of Part III we pass to normal coordinates, introducing the variables p_f and q_f by the relations

$$p_n = \left(\frac{M}{N}\right)^{\frac{1}{2}} \sum_f e^{ifn} p_f, \quad (47.9)$$

$$\xi_n = \frac{1}{(MN)^{\frac{1}{2}}} \sum_f e^{ifn} q_f. \quad (47.10)$$

Since p_n and ξ_n are real quantities, the following equalities hold:

$$p_{-f} = p_f^*, \quad q_{-f} = q_f^*. \quad (47.11)$$

Substituting (47.9) and (47.10) into (47.3) and taking into account (47.6), we find

$$H = H_0 + \frac{1}{2} \sum_f (p_f^2 + \omega_f^2 q_f^2). \quad (47.12)$$

The second term of (47.12) represents the Hamiltonian of a system of independent oscillators with frequencies ω_f . Transition to the quantum Hamiltonian is carried out, as usually, by replacing p_f and q_f by operators satisfying the commutation relations (26.2)–(26.4) of Part V.

Comparing expressions (47.5) and (47.12) with the corresponding formulae for the electromagnetic field operators (see § 100 of Part V) we satisfy ourselves of their formal identity.

As in the case of the electromagnetic field, it is convenient to introduce the creation operator \hat{b}_f^\dagger and the annihilation operator \hat{b}_f satisfying the com-

mutation relations (101.3) of Part V. Making use of the results of §101 of Part V, we can immediately write the quantized lattice energy

$$E = \sum_f (\hat{b}_f \hat{b}_f^\dagger + \frac{1}{2}) \hbar \omega_f = \sum_f (n_f + \frac{1}{2}) \hbar \omega_f. \quad (47.13)$$

Just as in the case of the electromagnetic field, a system of quantized waves can be looked upon as a system of independent quantum particles (bosons) called phonons. Formula (47.13) shows that the energy of a phonon with wave number f is equal to $\hbar \omega_f$.

In passing to a three-dimensional crystal the general situation becomes, in essence, only a little more complex. The displacement ξ_n can be written in the form of the vector

$$\xi_n = \frac{1}{(MN)^{\frac{1}{2}}} \sum_f \sum_{j=1}^3 \mathbf{e}_{fj} q_{fj} e^{i(\mathbf{f} \cdot \mathbf{n} - \omega_f t)}, \quad (47.14)$$

where \mathbf{e}_{fj} are the three polarization vectors: the longitudinal polarization vector \mathbf{e}_{f1} directed parallel to the vector \mathbf{f} , and the transverse polarization vectors \mathbf{e}_{f2} and \mathbf{e}_{f3} directed perpendicular to the vector \mathbf{f} . From the fact that ξ_n is real it follows that $q_{fj}^* = q_{-fj}$. The wave vector \mathbf{f} is defined to within the reciprocal lattice vector \mathbf{K} .

The Hamiltonian of a three-dimensional lattice is of the form

$$H = \frac{1}{2} \sum_{fj} (p_{fj}^* p_{fj} + \omega_{fj}^2 q_{fj}^* q_{fj}). \quad (47.15)$$

For a given value of \mathbf{f} there are three phonons with frequencies ω_{f1} , ω_{f2} and ω_{f3} .

In contrast to photons, for which there is a linear relation between the frequency ω and the wave number f , for phonons there is always a complex dispersion law.

For a linear chain the dispersion law is defined by formula (47.6); for a three-dimensional crystal there are different dispersion laws for different polarizations, i.e. $\omega_{fj} = \omega(\mathbf{f}, j)$.

Only for small values of the wave vectors in crystals with cubic symmetry can one set

$$\omega_{fj} = \omega_f = cf, \quad (47.6')$$

To carry out the quantization it is convenient to introduce the phonon creation operator $\hat{b}_{\mathbf{f}j}^{\dagger}$ and the phonon annihilation operator $\hat{b}_{\mathbf{f}j}$, setting

$$\hat{q}_{\mathbf{f}j} = \left(\frac{\hbar}{2\omega_{\mathbf{f}}} \right)^{\frac{1}{2}} \hat{b}_{\mathbf{f}j}; \quad \hat{q}_{\mathbf{f}j}^{\dagger} = \left(\frac{\hbar}{2\omega_{\mathbf{f}}} \right)^{\frac{1}{2}} \hat{b}_{\mathbf{f}j}^{\dagger}. \quad (47.16)$$

The commutation rules for the phonon creation and annihilation operators (101.3) of Part V must be written in the form

$$\hat{b}_{\mathbf{f}j} \hat{b}_{\mathbf{f}'j'}^{\dagger} - \hat{b}_{\mathbf{f}'j'}^{\dagger} \hat{b}_{\mathbf{f}j} = \delta_{\mathbf{f}\mathbf{f}'} \delta_{jj'}. \quad (47.17)$$

Then

$$E = \sum_{\mathbf{f}} \sum_j (\hat{b}_{\mathbf{f}j}^{\dagger} \hat{b}_{\mathbf{f}j} + \frac{1}{2}) \hbar \omega_{\mathbf{f}j} = \sum_{\mathbf{f}} \sum_j (n_{\mathbf{f}j} + \frac{1}{2}) \hbar \omega_{\mathbf{f}j}. \quad (47.18)$$

Generalization of the theory to the case of lattices with a basis does not introduce any essentially new points.

In the most general case it can be stated that the thermal excitation of the lattice is described by a system of elementary independent particles — phonons. Thus the description of collective excitations (waves) in a crystal lattice is very similar to that of the field of electromagnetic radiation in a cavity. However, it is necessary to stress that this analogy is to a certain degree formal. Whereas photons are as real as any other particles: electrons, mesons or protons, phonons are a fictitious, formally introduced concept. As a matter of fact the Hamiltonian in the form (47.3), allowing the introduction of normal coordinates and the transformation into the form (47.12), is not exact. It was obtained as a result of disregarding terms of the third and higher order of small quantities ξ_n in formula (47.2), i.e. in a harmonic approximation. Taking into account anharmonicity makes it impossible to bring the Hamiltonian to the form (47.3), i.e. to a sum of quadratic terms. Hence the possibility itself of introducing the concept of phonons is closely associated with the approximate consideration of the thermal motion of a crystal lattice. In contrast to true Bose particles (photons), phonons are said to be quasiparticles. The difference between quasiparticles (phonons) and true particles (photons) manifests itself particularly clearly by the fact that phonons cannot be assigned a definite value of momentum. Indeed, the momentum of a free particle, as well as its energy, runs over a continuous, in no way bounded sequence of values. The energy and momentum are conserved in any interaction. For the phonon, the role of momentum is formally played by the quantity $\hbar\mathbf{k}$.

However, the value of this quantity lies in the definite interval (47.8). Moreover, the value of the vector $\hbar\mathbf{f}$ is not determined unambiguously but only to within the quantity $\hbar\mathbf{K}$, where \mathbf{K} is the reciprocal lattice vector. Therefore the vector $\hbar\mathbf{f}$ is said to be a quasi-momentum. In contrast to true momentum, the value of the quasi-momentum need not be conserved in interactions of the phonon with other phonons or electrons. The crystal lattice can, in an arbitrary way, give to or take away from the phonon the quasi-momentum $\hbar\mathbf{K}$.

In spite of the fictitious character of the quasiparticles (phonons) their introduction has turned out to be very fruitful. It represents the basis for the consideration of all processes occurring in solids. Moreover, the introduction of elementary excitations as certain quasiparticles is an approach typical of modern physics in describing excited states in systems with a large number of particles.

In a number of problems, the excitation corresponding to the collective motion of a macroscopic system can be described by means of certain coordinates X_i so that the total Hamiltonian of the system, \hat{H} , turns out to be equal to

$$\hat{H} = \sum \hat{H}_i, \quad \text{where} \quad \hat{H}_i = ap_i^2 + f(X_i).$$

Then the energy of the system is also written in the form of a sum

$$E = \sum E_i.$$

It very often turns out to be possible to find variables for which the Hamiltonian \hat{H}_i has a form similar to that of the Hamiltonian for an oscillator. It is then said that the collective motion of the system can be decomposed into a set of elementary excitations. In this case the eigenvalues of the operator \hat{H}_i corresponding to the i th degree of freedom of the collective motion can be looked upon as the energy of a certain quasiparticle. The case discussed above, of the lattice vibrations of a crystal as a whole leads to one of the forms of the quasiparticles; phonons. Each quasiparticle corresponds to one degree of freedom of the collective motion of the system. The introduction of quasiparticles has the advantage of allowing the motion of a system of real interacting particles to be compared to a system of non-interacting or weakly interacting quasiparticles.

The energy spectrum of the set of quasiparticles is the same as that of the real system. The possibility of introducing quasiparticles occurred in a number of problems associated with the theory of systems of many particles and mainly in solid-state theory. Hence, at present there is in theoretical physics a

multitude of quasiparticles, for example polarons (electrons in polar crystals surrounded by a 'cloud' of phonons), excitons (electron-hole pairs formed in a semiconductor), magnons (elementary excitations in ferromagnetic solids), plasmons etc. These quasiparticles are not truly existing particles. However, their formal introduction reflects the character of the processes occurring in many-particle systems, and allows one to make use of a convenient and well elaborated mathematical apparatus.

The concept of phonons is basic in considering the motion of electrons in a crystal lattice. We shall see that the interaction of the electrons with the vibrating lattice and their scattering on the atoms of the lattice gives rise to electrical resistance. This interaction, as will be seen from the calculations, can formally be considered as a result of the collisions between electrons and phonons which, as a gas, fill the volume of the solid body. These collisions obey the laws of conservation of momentum and energy in the system of electrons and phonons.

For the concept of quasiparticles to make sense, it is necessary that their lifetime be sufficiently large. If $\bar{\epsilon}$ is the mean energy of a quasiparticle, and $\Delta\epsilon$ the uncertainty in its energy, then the following inequality must be fulfilled

$$\bar{\epsilon} \geq \Delta\epsilon \sim \hbar/\tau, \quad (47.19)$$

where τ is the lifetime of the quasiparticle.

The processes determining the lifetime of a quasiparticle can be any processes of absorption and scattering. In the case of phonons, the origins of absorption and scattering can be impurities in the crystal and the anharmonicity corresponding to collisions between phonons.

In a pure crystal the number of impurities is small. At low temperatures the anharmonicity is also small. Hence at low temperatures sufficiently far from the melting point the lifetime of phonons turns out to be very large and the concept of phonons as quasiparticles makes sense.

§48. Wave function of an electron moving in a periodic field

Before studying the motion of a system of electrons in an external field we shall consider the motion of an electron in the periodic field of a crystal lattice. In succeeding sections it will be shown how this solution can be used for investigating the entire system of electrons.

For an electron moving in a crystal lattice the Schrödinger equation is of

the form

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + U \right] \psi_{\text{el}} = E_n \psi_{\text{el}}. \quad (48.1)$$

In this equation the potential energy U is a periodic function with a period equal to the crystal lattice spacing. It satisfies the periodicity condition

$$U(\mathbf{r} + \mathbf{n}) = U(\mathbf{r}), \quad (48.2)$$

where the vector \mathbf{n} is defined by formula (46.1). We introduce the translation operator $\hat{T}_{\mathbf{n}}$ defined in the following way:

$$\hat{T}_{\mathbf{n}} \psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{n}). \quad (48.3)$$

Since the potential energy is a periodic function, it is obvious that the operator $\hat{T}_{\mathbf{n}}$ commutes with the Hamiltonian $\hat{H} = -(\hbar^2/2m) \nabla^2 + U$. Hence it follows that the wave functions of stationary states can be chosen in such a way as to be the eigenfunctions of the two operators \hat{H} and $\hat{T}_{\mathbf{n}}$. This means that the following equations must hold

$$\begin{aligned} \hat{T}_{\mathbf{n}} \psi(\mathbf{r}) &= \alpha_{\mathbf{n}} \psi(\mathbf{r}), \\ \hat{H} \psi(\mathbf{r}) &= E_n \psi(\mathbf{r}). \end{aligned} \quad (48.4)$$

Here $\alpha_{\mathbf{n}}$ is an eigenvalue of the operator $\hat{T}_{\mathbf{n}}$. It is easily shown that the values of $\alpha_{\mathbf{n}}$ are in modulus equal to one. Indeed, the probabilities of finding the electron at points \mathbf{r} and $\mathbf{r} + \mathbf{n}$ must be the same by virtue of the periodic properties of the lattice, i.e.

$$|\psi(\mathbf{r} + \mathbf{n})|^2 = |\psi(\mathbf{r})|^2.$$

Using (48.4), we obtain

$$|\alpha_{\mathbf{n}} \psi(\mathbf{r})|^2 = |\psi(\mathbf{r})|^2,$$

which proves our statement.

Without restricting the generality, the coefficient $\alpha_{\mathbf{n}}$ can be written in the form

$$\alpha_{\mathbf{n}} = e^{i\mathbf{k} \cdot \mathbf{n}}. \quad (48.5)$$

For reasons which will become clear from what follows, \mathbf{k} is said to be the effective wave vector of an electron moving in a periodic field. Below we shall for brevity frequently call \mathbf{k} the wave vector.

Consequently, for the eigenfunctions of the operators \hat{H} and $\hat{T}_{\mathbf{n}}$, in correspondence with (48.4), the following relation must hold:

$$\psi(\mathbf{r} + \mathbf{n}) = e^{i\mathbf{k} \cdot \mathbf{n}} \psi(\mathbf{r}) . \quad (48.6)$$

In the most general case, the function satisfying condition (48.6) can be written in the form

$$\psi(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) , \quad (48.7)$$

where $u_{\mathbf{k}}(\mathbf{r})$ is a periodic function with a period equal to the lattice period

$$u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{n}) . \quad (48.8)$$

Substituting the wave function ψ defined by (48.7) into the Schrödinger equation, we find the equation for $u_{\mathbf{k}}(\mathbf{r})$

$$\left\{ -\frac{\hbar^2}{2m} (\nabla + i\mathbf{k})^2 + (U - E) \right\} u_{\mathbf{k}}(\mathbf{r}) = 0 . \quad (48.9)$$

Eq. (48.9) shows that, besides other quantum numbers, the energy of the electron also depends on the wave vector \mathbf{k} .

If we wanted to take into account surface effects, we should require that the wave function decrease outside the crystal. At large distances from the crystal it must tend to zero. However, we shall not take up a study of surface effects here, but replace the boundary condition mentioned by a periodicity condition. For a cubic crystal with edge L it is of the form $e^{ik_i L} = 1$. This corresponds to the condition $\psi(x=0) = \psi(x=L)$ in the cyclic planes. Hence it follows that the vector components k_i take on values $k_i = (2\pi/L) l_i$, where l_i runs over a sequence of integer values.

However, taking into account that the dimensions L are very large, one can also approximately assume \mathbf{k} to be a continuously varying quantity under the requirement of cyclicity. Another feature of the vector \mathbf{k} is the ambiguity of its definition. If the reciprocal lattice vector \mathbf{K} is added to the vector \mathbf{k} , then in formula (48.6), defining \mathbf{k} , one can write

$$\psi(\mathbf{r} + \mathbf{n}) = e^{i(\mathbf{k} + \mathbf{K}) \cdot \mathbf{n}} \psi(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{n}} \psi(\mathbf{r}) ,$$

since, in correspondence with (48.2), $e^{i\mathbf{K}\cdot\mathbf{n}} = 1$. Thus the replacement of \mathbf{k} by $\mathbf{k}+\mathbf{K}$ has no effect on the definition of this vector. The wave vector \mathbf{k} is said to be defined to within the reciprocal lattice vector \mathbf{K} .

Formula (48.7) shows that the wave function of an electron moving in a periodic field is of the form of a modulated plane wave, i.e. a plane wave with amplitude $u_{\mathbf{k}}(\mathbf{r})$ varying in space. Thus there is an analogy between the wave function of a particle in a periodic field and the wave function of a free particle. This analogy justifies the term effective wave vector \mathbf{k} . At the same time it should be stressed that an electron moving in a periodic field is a bound and not a free particle. Its momentum has no definite constant value.

We shall frequently normalize $\psi(\mathbf{r})$ by the condition

$$\int_G |\psi(\mathbf{r})|^2 dV = N, \quad (48.10)$$

where the integral is taken over the so-called basic region of the crystal, i.e. over the region bounded by the nearest cyclic planes. N is the number of particles in the basic region of the crystal. If the wave function is written in the form

$$\psi = \frac{1}{(N)^{\frac{1}{2}}} e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \quad (48.11)$$

then (48.10) goes over into the following normalization condition:

$$\int_{\tau} |u_{\mathbf{k}}|^2 dV = 1. \quad (48.12)$$

The integration in this formula is carried out over a unit cell.

§49. The energy spectrum of an electron moving in a periodic field

Expression (48.7) for the wave function is of a general character. In order to get some idea of the concrete form of the function $u_{\mathbf{k}}$ and about the energy spectrum $E(\mathbf{k})$ of an electron in a crystal lattice, it is necessary to make certain assumptions about the character of the potential energy of an electron in a crystal. We shall assume that the wave function and the energy spectrum of an electron in a certain isolated j th atom are known, i.e. that the solution

of the following equation is known:

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + (U_j - E(n)) \right\} \varphi_j = 0, \quad (49.1)$$

where U_j is the potential energy of an electron in the j th atom, n is the quantum number, and $E(n)$ are the eigenvalues of the energy of the electron in the atom. If a crystal lattice is formed of atoms of a given type, then owing to the action of the fields of neighbouring atoms on the electron its wave function and energy levels will change. We shall, however, assume that this action is sufficiently weak, so that (1) the action of only the nearest neighbours need be taken into account, (2) the shift of energy levels due to the action of neighbouring atoms is small in comparison with the spacings between neighbouring energy levels of an isolated atom.

The solution of the Schrödinger equation for an electron in a crystal lattice obtained on the basis of these assumptions is said to be the solution in the tight binding approximation.

To characterize an electron in a crystal lattice consisting of N atoms in the zeroth order approximation, i.e. without taking into account the action of neighbours, one can write N wave functions satisfying the Schrödinger equation (49.1) for one and the same value of the energy. Therefore the state of the electron is N -fold degenerate (roughly speaking, the electron can be localized at any of the N atoms). The interaction of the electron with the neighbouring atoms removes this degeneracy.

As is done in the perturbation theory of degenerate states, we seek the wave function in the zeroth order approximation in the form

$$\psi = \sum_j C_j \varphi_j, \quad (49.2)$$

where C_j are constants, and φ_j is the solution of eq. (49.1) for the j th atom. The summation is carried out over all atoms of the crystal. We understand \mathbf{j} to be the radius vector going from the origin to the nucleus of the j th atom.

To simplify the calculations we shall assume that all the electrons are in the non-degenerate s-state.

We shall denote the energy of an electron in an atom by $E(s)$. Then the energy of an electron in the crystal will be of the form

$$E(\mathbf{k}, s) = E(s) + E^{(1)}(\mathbf{k}, s). \quad (49.3)$$

It is different from the energy of an electron in an atom and depends on the

wave vector \mathbf{k} . Substituting the wave function (49.2) and the energy (49.3) into the Schrödinger equation for an electron moving in a crystal, we have

$$\sum_j C_j \nabla^2 \varphi_j + \frac{2m}{\hbar^2} \left[E(s) + E^{(1)}(\mathbf{k}, s) - \sum_l U_l(\mathbf{r}) \right] \sum_j C_j \varphi_j = 0. \quad (49.4)$$

Here we have written the potential energy U of an electron in a crystal lattice in the form of the sum $\sum_l U_l(\mathbf{r})$, where $U_l(\mathbf{r})$ is the potential energy of an electron in the field of the l th atom and the summation is carried out over all atoms of the lattice.

Since the potential energy in eq. (49.4) cannot be divided into an unperturbed part and a perturbation, the solution cannot be carried out by means of ordinary perturbation theory. Our purpose is to find an approximate solution of eq. (49.4). The electron in an unperturbed state is considered to be localized at the j th atom. The interaction of the electron with the atoms of the lattice which are nearest to the j th atom is a small perturbation.

Let us write the major term of the potential energy of the electron, the energy of interaction with the j th atom, in the form

$$\sum_l U_l(\mathbf{r}) = \sum_l' U_l(\mathbf{r}) + U_j(\mathbf{r}) = U' + U_j(\mathbf{r}),$$

where the prime denotes that the term with $l = j$ is omitted in the sum. Then we can write

$$\sum_j C_j \left\{ \nabla^2 \varphi_j + \frac{2m}{\hbar^2} [E(s) - U_j] \varphi_j \right\} + \sum_j \left(E^{(1)}(\mathbf{k}, s) - \sum_l' U_l \right) C_j \varphi_j = 0.$$

By virtue of (49.1) the bracket reduces to zero, and we arrive at the equation

$$\sum_j C_j [E^{(1)}(\mathbf{k}, s) - U'] \varphi_j = 0. \quad (49.5)$$

By virtue of our assumption of the weak action of neighbouring atoms on the electron 'localized' at the j th atom, it can be assumed approximately that the wave functions φ do not overlap, i.e. we can set

$$\int \varphi_j \varphi_h^* dV = \begin{cases} 1 & \text{for } j = h, \\ 0 & \text{for } j \neq h. \end{cases}$$

Multiplying (49.5) by φ_h^* and integrating over the volume of the crystal, we have

$$\sum_j \int \varphi_h^* C_j (E^{(1)}(\mathbf{k}, s) - U') \varphi_j dV = 0.$$

Carrying out the integration, we obtain

$$C_h [E^{(1)}(\mathbf{k}, s) + \alpha] + \sum_{j \neq h} \beta(\mathbf{j} - \mathbf{h}) C_j = 0, \quad (49.6)$$

where the following notation is introduced:

$$\begin{aligned} \alpha &= - \int \varphi_h^* U' \varphi_h dV, \\ \beta(\mathbf{j} - \mathbf{h}) &= - \int \varphi_h^* U' \varphi_j dV. \end{aligned} \quad (49.7)$$

The integrals involved in α and β depend only on the difference between \mathbf{j} and \mathbf{h} but not on these quantities taken separately. Taking into account the small degree of overlap of the wave functions, we can confine ourselves in the sum of quantities $\beta(\mathbf{j} - \mathbf{h})$ to the terms corresponding to neighbouring atoms which are nearest to the \mathbf{h} th atoms

$$\sum_j \beta(\mathbf{j} - \mathbf{h}) C_j \simeq \sum_{\mathbf{n}_1} \beta(\mathbf{n}_1) C_{\mathbf{n}_1 + \mathbf{h}}.$$

In this sum the number of terms is equal to the number of nearest neighbours of the \mathbf{h} th atom. In the particular case of a cubic lattice, the number of nearest neighbours is 6. Furthermore, it is clear that in this case $\beta(\mathbf{n}_1) = \beta(-\mathbf{n}_1)$.

Giving \mathbf{h} all its possible values, we thus arrive at an infinite set of linear algebraic equations:

$$(E^{(1)}(\mathbf{k}, s) + \alpha) C_h + \sum_{\mathbf{n}} \beta(\mathbf{n}) C_{\mathbf{h} + \mathbf{n}} = 0. \quad (49.8)$$

We shall try to find the solution of this system in the form

$$C_h = e^{i\mathbf{k} \cdot \mathbf{h}}. \quad (49.9)$$

Substituting (49.9) into (49.8) and taking into account that an atom in a

cubic lattice has 6 nearest neighbours, we obtain

$$\begin{aligned} E^{(1)}(\mathbf{k}, s) &= -\alpha - \sum_{\mathbf{n}} \beta(\mathbf{n}) e^{i\mathbf{k} \cdot \mathbf{n}} = \\ &= -\alpha - 2\beta(\cos k_x a + \cos k_y a + \cos k_z a). \end{aligned} \quad (49.10)$$

Correspondingly, the wave function is of the form

$$\psi = \sum C_j \varphi_j = \sum e^{i\mathbf{k} \cdot \mathbf{j}} \varphi_j = e^{i\mathbf{k} \cdot \mathbf{r}} \sum e^{-i\mathbf{k} \cdot (\mathbf{r} - \mathbf{j})} \varphi_j.$$

For φ_j we can by definition write

$$\varphi_j(\mathbf{r}) = \varphi_s(\mathbf{r} - \mathbf{j}),$$

where the index s denotes that we consider an electron in the s -state. For the wave function ψ we have

$$\psi = e^{i\mathbf{k} \cdot \mathbf{r}} \sum e^{-i\mathbf{k} \cdot (\mathbf{r} - \mathbf{j})} \varphi_s(\mathbf{r} - \mathbf{j}). \quad (49.11)$$

Comparing with (48.7), we obtain

$$u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{j}} e^{-i\mathbf{k} \cdot (\mathbf{r} - \mathbf{j})} \varphi_s(\mathbf{r} - \mathbf{j}). \quad (49.12)$$

The function $u_{\mathbf{k}}$ possesses the needed properties of periodicity.

We now pass to a discussion of the expression for the electron energy defined by the quantities α and β . We note, first of all, that the wave function φ_j of an electron localized at a given atom \mathbf{j} has nodes only inside the atom. Outside the atom it decreases monotonically. We shall further show that the quantity U' involved in the expressions for α and β is negative.

Indeed, since the potential energy of an electron in a crystal lattice $U(\mathbf{r})$ and in an isolated atom U_j is negative (the electron is bound in the atom), one can write

$$U' = U(\mathbf{r}) - U_j = -|U(\mathbf{r})| + |U_j(\mathbf{r})|.$$

It is clear that the addition to the isolated atom of other atoms of the same nature (attracting the electron) leads to an additional increase in the binding of the electron in the system. Hence $|U(\mathbf{r})| > |U_j|$, which proves our state-

ment. From the monotonic character of φ_h and φ_j and the negative sign of the quantity U' it follows that the two integrals α and β defined by formulae (49.7) are essentially positive.

So, the addition to the energy of the s-state turns out to be negative; the electron is more strongly bound in the system of atoms than in the isolated atom. For small values of k one can expand $\cos ka$ in a series and write

$$E(\mathbf{k}, s) = E(s) - \alpha - 6\beta + \beta a^2 k^2 = E_{\min} + p^2/2m^* = E_{\min} + \hbar^2 k^2/2m^*, \quad (49.13)$$

where E_{\min} is a constant quantity independent of k , corresponding to a minimum value of energy

$$E_{\min} = E(s) - \alpha - 6\beta,$$

and m^* is equal to

$$m^* = \frac{\hbar^2}{2\beta a^2} = \frac{\hbar^2}{\partial^2 E / \partial k^2}. \quad (49.14)$$

Formula (49.13) shows that for small k the dependence of the electron energy on k is of the same form as for a free particle. Here the role of mass is played by the effective mass m^* , and the quantity $\mathbf{p} = \hbar\mathbf{k}$ is naturally called the momentum of the particle. However, it should be kept in mind that \mathbf{p} is not a real but an effective momentum. This is seen from the fact that the vector \mathbf{k} and along with it the vector \mathbf{p} are defined only to within the reciprocal lattice vector. Therefore the quantity \mathbf{p} is said to be a quasi-momentum. Just as the momentum conservation law holds for a free particle, the value of quasi-momentum is conserved for a bound particle in the periodic field of the crystal.

For large values of \mathbf{k} the expansion (49.13) loses its validity. When the following equalities are fulfilled

$$k_x a = \pi, \quad k_y a = \pi, \quad k_z a = \pi,$$

$\cos k_i a$ are equal to -1 and the energy of the electron takes on a maximum value

$$E\left(\frac{\pi}{a}; \frac{\pi}{a}; \frac{\pi}{a}; s\right) = E_{\max}.$$

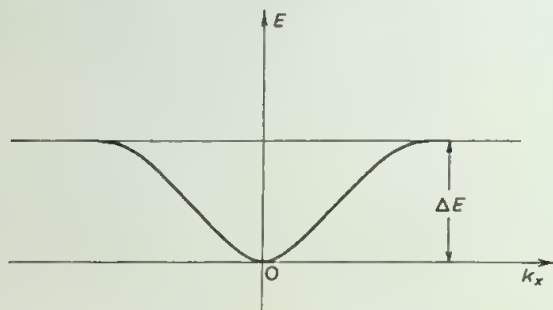


Fig. VI.7

The behaviour of the energy $E(\mathbf{k}, s)$ is shown schematically in fig. VI.7. Since $E(\mathbf{k}, s)$ depends on three variables, fig. VI.7 shows the dependence of the energy on k_x for fixed values of k_y and k_z .

We see that from the discrete energy level corresponding to the s-state of the electron in the atom a band of allowed energy values arises in the crystal. It should be stressed that the formation of the band of allowed energies is due to the interaction of the electron 'localized' at a given atom with the field of other atoms of the lattice. This interaction removes the degeneracy which exists in a system of N atoms and an electron. Since the electron can be 'localized' at any of the atoms, the multiplicity of the degeneracy is very large and the degenerate energy level, owing to the interaction mentioned, splits into a large number of sublevels. In that approximation in which the wave vector can be assumed to run over a continuous sequence of values (see §48), the set of sublevels can be considered as a continuous band of allowed energy values.

The allowed energy band arising from the ground energy level of the atom is called the lower energy band. The width of the lower energy band is equal to $\Delta E = E_{\max} - E_{\min} = 12\beta$. Exactly the same calculation can be carried out for states arising from the first excited state of an electron in an isolated atom. We shall assume it to be a p-state. The energy of the electron $E(\mathbf{k}, p)$ is then given by a formula of the same type as for $E(\mathbf{k}, s)$. However, it turns out* that one of the overlap integrals (integrals (49.7)) is negative, while the other two are positive.

Thus the first excited energy band of an electron moving in a crystal lat-

* See A.H.Wilson, *The theory of metals* (Cambridge University Press, Cambridge, 1953).

tice arises from the energy level corresponding to the p-state in an isolated atom. Subsequent energy levels of an isolated atom turn into corresponding bands in the crystal.

In complete accordance with assumption (2) (at the beginning of the section), the condition for qualitative applicability of the calculation given is the requirement that the width of the bands be small in comparison with the spacing between energy levels in the unperturbed atom. Since the width of bands is defined by integrals of the type (49.7), this requirement is equivalent to small overlap of the wave functions of the electron in neighbouring atoms

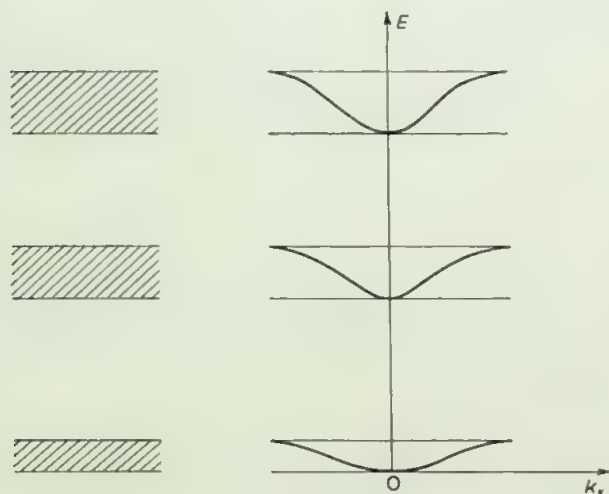


Fig. VI.8

The energy spectrum in a crystal in the tight binding approximation has the character of relatively narrow bands of allowed energies separated by wide bands of forbidden energies (fig. VI.8). This approximation holds, in fact, for relatively deep atomic electrons. For outer electrons the wave functions of neighbouring atoms are so strongly overlapped that the tight binding approximation is not valid. In other words, the probability of finding the electron near any point of the crystal lattice has almost the same value.

In this case it again turns out to be possible to carry out the calculation of the energy spectrum of the electron in the approximation of nearly free electrons. Namely, in the zero order approximation we do not take into account the action of the periodic field of the crystal lattice at all, and consider the

electron to be free. Its normalized wave function and its energy are of the form

$$\psi_0 = V^{-1/2} e^{i\mathbf{k} \cdot \mathbf{r}}, \quad E^0(\mathbf{k}) = \frac{\hbar^2 k^2}{2m}.$$

Here unperturbed states are, as a rule, degenerate. This is particularly obvious in the example of a cubic crystal. As a matter of fact, in correspondence with the foregoing (see §48), the vector components k_x, k_y, k_z can take on the following values: $k_i = (2\pi/a)l_i$ where l_i takes on integer values.

Thus if there is a vector \mathbf{k} directed along the x -axis and equal to $2\pi l_0/a$, where l_0 is a certain integer, then there are a number of other vectors having the same magnitude but differing in their directions. To find the perturbed wave function, we have to make use of perturbation theory for degenerate states. We set

$$\psi = \sum_{\mathbf{k}} V^{-1/2} C_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{r}}, \quad (49.15)$$

where the summation is carried out over all vectors \mathbf{k} whose ends lie on the sphere $|\mathbf{k}| = \text{const}$. To determine the first order correction to the energy according to perturbation theory we have from formula (54.4) of Part V

$$|U_{\mathbf{k}, \mathbf{k}'} - E^{(1)}\delta_{\mathbf{k}, \mathbf{k}'}| = 0, \quad (49.16)$$

where the matrix element $U_{\mathbf{k}, \mathbf{k}'}$ is equal to

$$U_{\mathbf{k}, \mathbf{k}'} = \frac{1}{V} \int e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r}} U(\mathbf{r}) dV.$$

Let us at first suppose that among the vectors \mathbf{k} over which the summation in (49.15) is carried out and whose ends lie at the surface of the sphere $|\mathbf{k}| = \text{const}$, there are no vectors satisfying the relation

$$\mathbf{k}' - \mathbf{k} = \mathbf{K}. \quad (49.17)$$

In this case all integrals $U_{\mathbf{k}, \mathbf{k}'}$ are small except that of the diagonal matrix element $U_{\mathbf{k}, \mathbf{k}}$. This can be demonstrated in the following way. The matrix element

$$U_{\mathbf{k}, \mathbf{k}'} = \frac{1}{V} \int U e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r}} dV$$

is determined by integrating over all the space of the crystal. It is convenient to pass to integration over a unit cell

$$U_{\mathbf{k}, \mathbf{k}'} = \frac{1}{V} \sum_{\text{over cells}} \int_{\tau_n} U e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r}} dV.$$

We have to transform the integral over the n th unit cell, introducing the variables $\mathbf{r} = \mathbf{n} + \mathbf{r}'$, where \mathbf{n} is the vector directed from the origin to the n th atom of the cell. Then we obtain

$$\int_{\tau_n} U(\mathbf{r}) e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r}} dV = e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{n}} \int_{\tau_n} e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r}'} U(\mathbf{r}' + \mathbf{n}) dV'.$$

Making use of the property of periodicity of the function U , we find that the right-hand side of the last equality is of the form

$$e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{n}} \int_{\tau_0} U(\mathbf{r}') e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r}'} dV'.$$

For the matrix element $U_{\mathbf{k}, \mathbf{k}'}$ we have

$$\begin{aligned} U_{\mathbf{k}, \mathbf{k}'} &= \frac{1}{V} \sum_n e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{n}} \int_{\tau_0} U(\mathbf{r}) e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r}} dV = \\ &= \frac{1}{V} \int_{\tau_0} U(\mathbf{r}) e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{r}} dV \sum_n e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{n}}. \end{aligned} \quad (49.18)$$

In this formula the integration is carried out over a unit cell, while the summation is carried out over all lattice points.

The sum $B = \sum_n e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{n}}$ for arbitrary $\mathbf{k}' - \mathbf{k}$ is equal to

$$B = \sum_n e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{n}} = \begin{cases} N, & \text{if } \mathbf{k}' - \mathbf{k} = \mathbf{K}, \\ 0, & \text{if } \mathbf{k}' - \mathbf{k} \neq \mathbf{K}, \end{cases} \quad (49.19)$$

where N is the number of lattice points.

The upper equality (49.19) is obtained at once after making use of formula (46.2). The lower relation can be obtained if \mathbf{n} is replaced by $\mathbf{n} + \boldsymbol{\alpha}$,

where α is any vector of the lattice. For large values of N the quantity B cannot change when this substitution is made, since such a replacement is equivalent to a displacement of all lattice points

$$B = e^{i(\mathbf{k}' - \mathbf{k}) \cdot \alpha} \sum_{\mathbf{n}} e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{n}} = \sum_{\mathbf{n}} e^{i(\mathbf{k}' - \mathbf{k}) \cdot \mathbf{n}}.$$

If $\mathbf{k}' - \mathbf{k} \neq \mathbf{K}$, then it is obvious that $B = 0$.

We now see that the only diagonal matrix element differing from zero in (49.16) is equal to

$$U_{\mathbf{k}, \mathbf{k}} = \frac{1}{V} \int U(r) dV = \bar{U}.$$

The quantity \bar{U} represents the mean potential in the space of the lattice.

Evaluating determinant (49.16) and solving the equation leads to the unique root $E^{(1)} = \bar{U}$. In correspondence with eq. (54.3) of Part V, the coefficients $C_{\mathbf{k}}$ can be chosen in the following way:

$$\begin{aligned} 1) \quad C_{k_1} &= 1, & C_{k_2} &= C_{k_3} = \dots = 0, \\ 2) \quad C_{k_2} &= 1, & C_{k_1} &= C_{k_3} = \dots = 0 \text{ and so on.} \end{aligned}$$

We can now write the expressions for the wave function and energy. They are of the form

$$\psi_0(\mathbf{k}) = V^{-\frac{1}{2}} e^{i\mathbf{k} \cdot \mathbf{r}}, \quad E(\mathbf{k}) = E^0(\mathbf{k}) + U_{\mathbf{k}, \mathbf{k}} = E^0(\mathbf{k}) + \bar{U}.$$

The formulae found are in the same form as the results which would have been obtained if we had straight away used the theory of perturbations without degeneracy. This is due to the fact that the solution of eq. (49.16) is the unique root $E^{(1)} = \bar{U}$.

The correction can be obtained using that approximation. We make use of the usual expression of non-degenerate perturbation theory

$$E(\mathbf{k}) = E^0(\mathbf{k}) + \bar{U} + \sum_{\mathbf{k}'}' \frac{|U_{\mathbf{k}, \mathbf{k}'}|^2}{E_{\mathbf{k}} - E_{\mathbf{k}'}}. \quad (49.20)$$

In correspondence with (49.18) and (49.19), the matrix elements are differ-

ent from zero only if $\mathbf{k} - \mathbf{k}' = \mathbf{K}$. Expression (49.20) can now be written in the form

$$E(\mathbf{k}) = E^0(\mathbf{k}) + \bar{U} - \sum_{\mathbf{K}} \frac{|U_{\mathbf{k}, \mathbf{k}+\mathbf{K}}|}{(\hbar^2(2m)(\mathbf{K}^2 - 2\mathbf{k} \cdot \mathbf{K}))}. \quad (49.21)$$

The third term of formula (49.21) is a small correction, so that the spacing between the nearest levels is for practical purposes defined by the formula for free electrons. If, however, in one of the terms of the sum in formula (49.21) the denominator becomes zero, i.e. if for a certain \mathbf{K} the following equality holds

$$k'^2 - k^2 = \mathbf{K}^2 - 2\mathbf{k} \cdot \mathbf{K} = 0, \quad (49.22)$$

then this formula makes no sense. The correction to the energy $E(\mathbf{k})$ turns out to be considerable, and the use of perturbation theory is inadmissible. The presence of large terms in (49.21) allows one to assume the existence of large spacings between neighbouring levels.

Relation (49.22) means that for certain vectors \mathbf{k} and \mathbf{k}' the following equality is fulfilled

$$\mathbf{k}' - \mathbf{k} = \mathbf{K}, \quad |\mathbf{k}'| = |\mathbf{k}|. \quad (49.22')$$

Suppose that this equality holds only for one value of \mathbf{K} , i.e. for a definite value of \mathbf{k} and \mathbf{k}' . This means that the states of the electron have almost equal energies $E(\mathbf{k})$ and $E(\mathbf{k} + \mathbf{K})$, i.e. are two-fold degenerate. In this case, to find the correction to the energy, one has to make use of perturbation theory for two-fold degenerate levels. Thus eq. (49.16) is now written in the form

$$\begin{vmatrix} \bar{U} - E^{(1)}, & U_{\mathbf{k}, \mathbf{k}'} \\ U_{\mathbf{k}, \mathbf{k}'}, & \bar{U} - E^{(1)} \end{vmatrix} = 0.$$

Evaluating the determinant and solving the equation, we find for the correction to the energy

$$E^{(1)} = \bar{U} \pm U_{\mathbf{k}, \mathbf{k}'}.$$

This means that the energy undergoes a discontinuity and that the spacing between neighbouring levels is equal to $2|U_{\mathbf{k}, \mathbf{k}'}|$. It is obvious that the dis-

continuity in the energy occurs when equalities (49.22') are simultaneously fulfilled. Formulae (49.22') express the well-known Bragg reflection conditions. They single out in the crystal a definite direction of selective reflection of incident rays. Indeed, multiplying the first equality (49.22') by the lattice vectors a , b , c , we see that formulae (49.22') and (36.18) of Part IV are identical.

Thus, for a cubic crystal, in the approximation of nearly free electrons the energy as a function of the wave vector has the following behaviour. It increases with increasing k according to (49.21). When k reaches the values $k_x = k_y = k_z = \pm \pi/a$, Bragg reflection occurs. The energy undergoes a discontinuity, the width of the forbidden zone being equal to $2\bar{U}$. After this the energy is again expressed by formula (49.21). For $k_x = k_y = k_z = \pm 2\pi/a$ the energy again undergoes a discontinuity, and so on.

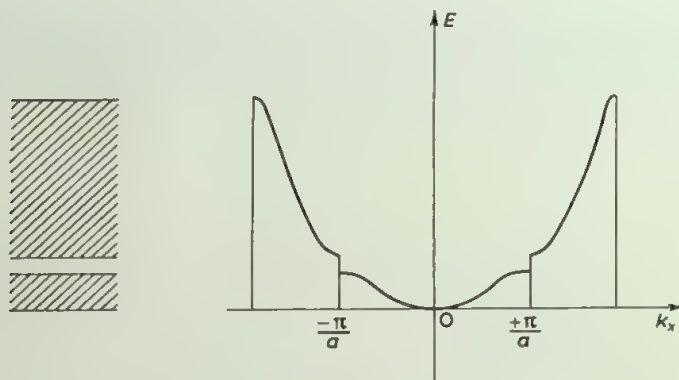


Fig. VI.9

Fig. VI.9 shows schematically the energy spectrum of a nearly free electron. The qualitative picture of the energy spectrum of a nearly free electron is the same as that of a strongly bound electron. A difference lies in the fact that in the first case broad allowed zones and narrow bands of forbidden energies are obtained, in contrast to narrow allowed zones and broad bands of forbidden energies in the second case. In real crystals the limiting cases of tight binding and nearly free electrons as well as the whole range of intermediate cases are found. However, in all cases the energy spectrum of an electron in the crystal has the character of allowed energy zones separated by regions of forbidden energy.

The quantitative calculation of the form of the dependence of $E(k)$ on k

for real crystals is a complex problem. A number of approximate methods have been devised for the calculation of the bands. Their description may be found in the specialist literature*.

In the next section it will be shown that the general result obtained in the one-particle model, i.e. for the case of the motion of one electron in a crystal lattice, also remains valid in the case of a crystal containing a large number of electrons. Thus the energy spectrum of electrons in a real crystal has the character of bands**.

The motion of an electron in a metal lattice or a semiconductor lattice can be characterized by its mean (group) velocity, which can be found by the general formula

$$\bar{\mathbf{v}} = \frac{\partial \omega}{\partial \mathbf{k}} = \frac{\partial E(\mathbf{k})}{\partial (\hbar \mathbf{k})}.$$

If the energy of a quasi-free electron is substituted for $E(\mathbf{k})$, then we arrive at the formula $\bar{v}_x = \hbar k_x / m$. In the case of a strongly bound electron the mean velocity is of the form $v_x = (2\beta a / \hbar) \sin k_x a$.

We see that the mean velocity of the electron differs from zero in both cases. A strongly bound electron moves with a mean velocity which is smaller, the smaller β , i.e. the smaller the width of the band.

We now assume that the crystal is placed in an external electric field of strength \mathcal{E} . The energy of an electron in the crystal will be equal to

$$E = E(\mathbf{k}) - e \mathcal{E} \cdot \mathbf{r}.$$

The mean acceleration of the electron

$$\bar{\mathbf{p}} = - \frac{\partial E}{\partial \mathbf{r}} = e \mathcal{E}$$

turns out to be the same as that of a free particle. The meaning of this result is easily understood if the expression for energy balance is set up. The mean rate at which work is done by the field on the electron is equal to

* A.H.Wilson, *The theory of metals* (Cambridge University Press, Cambridge, 1953); R. Seitz, *The modern theory of solids* (McGraw-Hill, New York, 1940); J.M.Ziman, *Principles of the theory of solids* (Cambridge University Press, Cambridge, 1964); H.Jones, *The theory of Brillouin zones and electronic states in crystals* (North-Holland, Amsterdam, 1960).

** We shall dwell neither on certain essential difficulties of the band theory of crystals nor on certain general limits on its applicability, particularly to polar crystals.

$$e\bar{v} \cdot \mathcal{E} = \frac{e}{\hbar} \frac{\partial E}{\partial \mathbf{k}} \cdot \mathcal{E}.$$

On the other hand, this work only goes into an increase of the energy of the electron

$$\frac{dE}{dt} = \frac{\partial E(\mathbf{k})}{\partial \mathbf{k}} \cdot \frac{d\mathbf{k}}{dt}.$$

Equating the two expressions for the rate at which work is done, we arrive at the expression

$$\frac{d\mathbf{k}}{dt} = \frac{e}{\hbar} \mathcal{E}.$$

Here we assume that the increase of energy of the electron takes place within the limits of a given band. The absence of energy loss for the electron leads to its acceleration by the field. In other words, the electron moves in an ideal lattice without any resistance. In §55 it will be shown that the basic origins of resistance are in the thermal vibrations of the lattice, disturbing the regularity of the arrangement of atoms on the lattice points.

In considering the motion of electrons in a periodic field we have disregarded the size of the ions themselves, assuming them to be crystal lattice points. At first sight it seems surprising that such a rough approximation and the model of nearly free electrons resulting from it is a good approximation to reality. The size of ions is not very small as compared with the spacing between them, the radius of an ion usually amounting to about one half of the lattice spacing.

In the region of space occupied by the ion the wave function of the electron must behave like the wave function inside the atom. The electron inside the atom has a large kinetic energy and, correspondingly, a small wavelength. Hence its wave function undergoes oscillations.

It is clear that such a behaviour differs very much from the smooth behaviour of the wave function of a nearly free electron. It turns out, however, that if we wish to describe the behaviour of the wave function of an electron in a region outside the ion and are not interested in its detailed behaviour over all space, we can make use of a method similar to that applied in scattering theory for small energies. We have seen in §93 of Part V that in this case the behaviour of a particle outside the region where its potential is large (outside the nucleus or atom) does not depend on its behaviour inside this region. Therefore the true wave function, which is the solution of the Schrödinger equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + U\right)\psi = E\psi, \quad (49.23)$$

can be replaced by the fictitious wave function φ satisfying the equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + U_{\text{ps}}\right)\varphi = E\varphi \quad (49.24)$$

with the same energy and a changed potential energy. The operator U_{ps} is called the pseudopotential. If U_{ps} is constructed in such a way that outside the region $r < R$ it is the same as U , $U_{\text{ps}} = U$ for $r \geq R$, then outside the ion the functions ψ and φ satisfy one and the same Schrödinger equation. By choosing the pseudopotential U_{ps} properly, one can make the fictitious wave function φ extrapolate smoothly into the region $r < R$. The smoothed wave function φ will describe the motion of the electron in the crystal. Outside the ion it is the same as ψ , and the action of the field on φ in the region $r < R$ is a weak perturbation.

The choice of the pseudopotential is not quite an unambiguous procedure. Correspondingly, φ may also be of a somewhat different form. For φ , use is often made of the expression

$$\varphi = \sum a_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} - \sum b_{\mathbf{k}'} \phi_{\mathbf{k}'}(\mathbf{r}), \quad (49.25)$$

where $\phi_{\mathbf{k}'}(\mathbf{r})$ are wave functions for the internal states of the electron in the atom. These functions are strongly localized, and for $r > R$ actually reduce to zero. Hence here φ is the same as the wave function of a quasi-free particle. The coefficients can be chosen from the condition of orthogonality of φ and $\phi_{\mathbf{k}'}$.

On substituting (49.25) into the Schrödinger equation, one can choose the expression for the pseudopotential U_{ps} .

In the light of the above it becomes clear why the behaviour of electrons in a lattice is sufficiently well described by the model of nearly free electrons.

In the cases where one deals with the properties of electrons associated with their motion in the crystal as a whole, the behaviour of the electrons can be described by the fictitious wave function φ . This function in regions $r < R$ only slightly depends on the weak field of the crystal lattice whereas for $r > R$ it is the same as the true wave function. On the whole, φ displays a smooth behaviour, changing at distances of the order of the lattice constant, as is to be expected for quasi-free particles.

§50. A system of electrons in a solid

We can now go on to a discussion of the properties of a system of electrons in a solid. As we shall see below, the number of non-localized electrons moving throughout the volume of a crystal can vary over a very wide range. For metals it is very large, amounting to about 10^{23} electrons per cm^3 . For such high densities the criterion of degeneracy of the electron gas (see §80 of Part III) $V(2\pi mkT)^{3/2}/N(2\pi\hbar)^3 \ll 1$ turns out to be fulfilled up to temperatures of the order of 4–5 thousand degrees. This means that for such a high density the electrons always form a degenerate Fermi gas.

If, however, the kinetic energy of a degenerate Fermi gas is compared with its Coulomb interaction energy, then, as was shown in §79 of Part III, we have the relation

$$\frac{\epsilon_{\max}}{(e^2/r)} = \frac{\hbar^2}{2m} \left(\frac{3}{8\pi} \frac{N}{V} \right)^{2/3} \frac{e^2}{(V/N)^{1/3}} \simeq \left(\frac{N}{V} \right)^{1/3} \left(\frac{\hbar^2}{e^2 m} \right) \sim 1$$

for $N/V \sim 10^{23}$.

Thus the Coulomb interaction between electrons in a metal is strong, and one should speak of an electron fluid rather than of an electron gas filling the metal. However, since this fluid is formed by electrons moving on a background of positively charged ions, as a matter of fact it is more correct to speak of a plasma filling the volume of the solid body. The most important difference between the plasma of a solid and a gas plasma is the fact that the former is degenerate and that quantum effects play an important role in it. Nevertheless, as we shall now see, the most important properties of a degenerate quantum plasma and a classical plasma turn out to be similar. This at least applies to the properties of screening and the existence of plasma waves. In order not to complicate the calculations, we shall confine ourselves to the approximation in which the charge of positive ions can be considered to be uniformly 'spread' throughout the crystal and to form a positive background. Taking into account the discrete distribution of positive charges does not introduce any essential change in the result obtained.

Let us consider a system of free electrons, with the distribution function $f(\epsilon)$, and a positive background in a certain volume acted upon by an external variable field. We shall assume that the applied external field produces in the system the potential

$$\varphi = \lim_{\delta \rightarrow 0} \int \int \varphi(\mathbf{q}, \omega) e^{i(\mathbf{q} \cdot \mathbf{r} - \omega t)} e^{\delta t} d\mathbf{q} d\omega. \quad (50.1)$$

Here $\varphi(\mathbf{q}, \omega)$ represents the Fourier component. The factor $e^{\delta t}$ means that the field is adiabatically (slowly) applied during the time interval from $t \rightarrow -\infty$ to $t = 0$. For such a case the field will not give rise to transitions in the system up to the instant of time $t = 0$. For $t > 0$, transitions will occur in the system or, in other words, a response to the applied perturbation will arise. This response is conveniently characterized by the dielectric constant $\epsilon(\mathbf{q}, \omega)$. To calculate $\epsilon(\mathbf{q}, \omega)$ it is convenient to express it in terms of the density of the charge induced in the system by the external field.

Using formula (31.12) of Part IV, the obvious expressions for Ohm's law and the continuity equation in Fourier components

$$\mathbf{j}(\mathbf{q}, \omega) = \sigma(\mathbf{q}, \omega) \mathbf{E}(\mathbf{q}, \omega), \quad i\mathbf{q} \cdot \mathbf{j}(\mathbf{q}, \omega) = i\omega \rho(\mathbf{q}, \omega),$$

and the relation $\mathbf{E}(\mathbf{q}, \omega) = -i\mathbf{q}\varphi(\mathbf{q}, \omega)$, where $\varphi(\mathbf{q}, \omega)$ is the Fourier component of the scalar potential, we obtain for dielectric permeability

$$\begin{aligned} \epsilon(\mathbf{q}, \omega) &= 1 + \frac{4\pi i \sigma(\mathbf{q}, \omega)}{\omega} = \\ &= 1 + \frac{4\pi i \mathbf{j} \cdot \mathbf{E}}{\omega E^2} = 1 + \frac{4\pi i}{q^2} \frac{\mathbf{q} \cdot \mathbf{E}}{E^2} \rho(\mathbf{q}, \omega) = 1 - \frac{4\pi \rho(\mathbf{q}, \omega)}{q^2 \varphi(\mathbf{q}, \omega)}. \end{aligned} \quad (50.2)$$

Our next problem is to calculate the Fourier component of the charge density $\rho(\mathbf{q}, \omega)^*$.

With this end in view, we shall consider a free electron in the initial state with wave vector \mathbf{k} and energy $\epsilon(\mathbf{k})$.

The wave function of an electron in an unperturbed state can be written in the form

$$\psi_{\mathbf{k}} = V^{-\frac{1}{2}} e^{i\mathbf{k} \cdot \mathbf{r}} e^{i\epsilon(\mathbf{k})t/\hbar}.$$

Let us consider the perturbation produced by an external field of frequency ω , i.e. let us assume that the perturbed Hamiltonian is of the form

$$\hat{H}' = e\varphi(\mathbf{q}, \omega) e^{\delta t} (e^{i(\mathbf{q} \cdot \mathbf{r} - \omega t)} + e^{i(\mathbf{q} \cdot \mathbf{r} + \omega t)}).$$

* J.H.Ziman, *Principles of the theory of solids* (Cambridge University Press, Cambridge, 1964) ch. 5.

For a field applied in such a way, no transitions will be caused in the system for times up to $t = 0$. For $t > 0$, under the action of the perturbation the free electron assumes the wave function

$$\psi(t) = \psi_{\mathbf{k}} + c_1(t) \psi_{\mathbf{k}+\mathbf{q}} + c_2(t) \psi_{\mathbf{k}-\mathbf{q}}. \quad (50.3)$$

In first order of perturbation theory (see §55 of Part V), we have for c_1 and c_2

$$c_1(t) = \frac{1}{i\hbar} \int_{-\infty}^t (\hat{H}')_{\mathbf{k}, \mathbf{k}+\mathbf{q}}^* e^{-i[\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}+\mathbf{q})]t/\hbar} dt.$$

The matrix element is taken with respect to the time-independent electron gas wave functions. It is equal to

$$\begin{aligned} (\hat{H}')_{\mathbf{k}, \mathbf{k}+\mathbf{q}}^* &= \\ &= e\varphi(\mathbf{q}, \omega) \left[\int e^{-i(\mathbf{k}+\mathbf{q})\cdot\mathbf{r}} e^{i(\mathbf{q}\cdot\mathbf{r} - \omega t)} e^{\delta t} e^{i\mathbf{k}\cdot\mathbf{r}} dV + \dots \right] = e\varphi(\mathbf{q}, \omega) e^{-i\omega t}. \end{aligned}$$

Analogously

$$(H')_{\mathbf{k}, \mathbf{k}-\mathbf{q}} = e\varphi(\mathbf{q}, \omega) e^{i\omega t}.$$

Thus

$$c_1(t) = \frac{e\varphi(\mathbf{q}, \omega)}{i\hbar} \left(\int_{-\infty}^t e^{\delta t} e^{i\{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}+\mathbf{q}) + \hbar\omega\}t/\hbar} dt \right).$$

The factor $e^{\delta t}$ ensures the convergence of the integrals at the lower limit. Carrying out the integration, and then passing to the limit $\delta \rightarrow 0$, we finally find for $c_1(t)$

$$c_1(t) = - \frac{e\varphi(\mathbf{q}, \omega) e^{i\{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}+\mathbf{q}) + \hbar\omega\}t/\hbar}}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} + \mathbf{q}) + \hbar\omega}. \quad (50.4)$$

Analogously

$$c_2(t) = \frac{e\varphi(\mathbf{q}, \omega) e^{i\{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}-\mathbf{q}) - \hbar\omega\}t/\hbar}}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} - \mathbf{q}) - \hbar\omega}. \quad (50.5)$$

Let us now find the electron density induced by a perturbation in a system of free electrons. It is obvious that for one electron the induced charge density can be written in the form

$$\begin{aligned}\Delta\rho &= e \{ |\psi(t)|^2 - |\psi_k|^2 \} \simeq \\ &\simeq e \{ c_1^* \psi_{\mathbf{k}+\mathbf{q}}^* \psi_{\mathbf{k}} + c_2^* \psi_{\mathbf{k}-\mathbf{q}}^* \psi_{\mathbf{k}} + c_1 \psi_{\mathbf{k}+\mathbf{q}} \psi_{\mathbf{k}}^* + c_2 \psi_{\mathbf{k}-\mathbf{q}} \psi_{\mathbf{k}}^* \}.\end{aligned}$$

Here we have dropped the terms of the second order of small quantities c_1^2 and c_2^2 . Substituting the values of c_1 and c_2 from (50.4) and (50.5), we find

$$\begin{aligned}\Delta\rho &= -e^2 \varphi(\mathbf{q}, \omega) \left\{ \frac{e^{-i(\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}+\mathbf{q}) + \hbar\omega)t/\hbar}}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}+\mathbf{q}) + \hbar\omega} \times \right. \\ &\quad \times e^{-i(\mathbf{k}+\mathbf{q})\cdot\mathbf{r}} e^{i\mathbf{k}\cdot\mathbf{r}} e^{i\epsilon(\mathbf{k})t/\hbar} e^{-i\epsilon(\mathbf{k}+\mathbf{q})t/\hbar} + \dots \left. \right\} = \\ &= e^2 \varphi(\mathbf{q}, \omega) (e^{-i(\mathbf{q}\cdot\mathbf{r} + \omega t)} + e^{-i(\mathbf{q}\cdot\mathbf{r} - \omega t)}) \times \\ &\quad \times \frac{1}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}+\mathbf{q}) + \hbar\omega} + \frac{1}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}-\mathbf{q}) - \hbar\omega}.\end{aligned}\quad (50.6)$$

The total charge density induced in a system of electrons by an external field is equal to

$$\begin{aligned}\bar{\rho}(\mathbf{q}, \omega) &= \int \Delta\rho f(\mathbf{k}) d\mathbf{k} = -e^2 \varphi(\mathbf{q}, \omega) (e^{i(\mathbf{q}\cdot\mathbf{r} - \omega t)} + e^{i(\mathbf{q}\cdot\mathbf{r} + \omega t)}) \times \\ &\quad \times \left\{ \int \frac{f(\mathbf{k}) d\mathbf{k}}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}+\mathbf{q}) + \hbar\omega} + \int \frac{f(\mathbf{k}) d\mathbf{k}}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}-\mathbf{q}) - \hbar\omega} \right\} = \\ &= -e^2 \varphi(\mathbf{q}, \omega) (e^{i(\mathbf{q}\cdot\mathbf{r} - \omega t)} + e^{i(\mathbf{q}\cdot\mathbf{r} + \omega t)}) \times \\ &\quad \times \left[\int \frac{f(\mathbf{k}) d\mathbf{k}}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}+\mathbf{q}) + \hbar\omega} + \int \frac{f(\mathbf{k}' + \mathbf{q}) d\mathbf{k}'}{\epsilon(\mathbf{k}' + \mathbf{q}) - \epsilon(\mathbf{k}) - \hbar\omega} \right] = \\ &= -e^2 \varphi(\mathbf{q}, \omega) (e^{i(\mathbf{q}\cdot\mathbf{r} - \omega t)} + e^{i(\mathbf{q}\cdot\mathbf{r} + \omega t)}) \int \frac{f(\mathbf{k}) - f(\mathbf{k}+\mathbf{q})}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k}+\mathbf{q}) + \hbar\omega} d\mathbf{k}.\end{aligned}\quad (50.7)$$

In the second integral we have made the replacement of variables $\mathbf{k} \rightarrow \mathbf{k}' + \mathbf{q}$

and then put $\mathbf{k}' \rightarrow \mathbf{k}$. Substituting this value of $\bar{\rho}$ into (50.2), we find

$$\epsilon(\mathbf{q}, \omega) = 1 + \frac{4\pi e^2}{q^2} \int \frac{f(\mathbf{k}) - f(\mathbf{k} + \mathbf{q})}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} + \mathbf{q}) + \hbar\omega} d\mathbf{k}. \quad (50.8)$$

The dielectric constant represents a basic characteristic of matter. Using formula (50.8), one can find the response of the system to an arbitrary perturbation of the plasma.

Let us examine the limiting cases $\omega \rightarrow 0$ (constant field) and $\hbar\omega \gg E(\mathbf{k}) - E(\mathbf{k} + \mathbf{q})$ (high-frequency field). For $\omega \rightarrow 0$

$$\epsilon(\mathbf{q}, 0) \simeq 1 + \frac{4\pi e^2}{q^2} \int \frac{f(\mathbf{k}) - f(\mathbf{k} + \mathbf{q})}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} + \mathbf{q})} d\mathbf{k}. \quad (50.9)$$

Here we shall confine ourselves to the case of long wavelengths $q \ll k$. Then we find

$$f(\mathbf{k}) - f(\mathbf{k} + \mathbf{q}) \simeq -\frac{\partial f}{\partial \epsilon} \frac{\partial \epsilon}{\partial \mathbf{k}} \cdot \mathbf{q}, \quad (50.10)$$

$$\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} + \mathbf{q}) \simeq -\frac{\partial \epsilon}{\partial \mathbf{k}} \cdot \mathbf{q}.$$

Hence

$$\epsilon(\mathbf{q}, 0) = 1 + \frac{4\pi e^2}{q^2} \int \left(\frac{\partial f}{\partial \epsilon} \right) \rho(\epsilon) d\epsilon = 1 + \frac{4\pi e^2}{q^2} \rho(\epsilon_F), \quad (50.11)$$

where $\rho(\epsilon_F)$ is the charge density at the Fermi surface,

$$\rho(\epsilon_F) = \frac{3N}{2\epsilon_F}. \quad (50.12)$$

Here we have made use of the property of the Fermi distribution $f(\epsilon)$, which for a low temperature has the form of a step function (see §80 of Part III).

For values of \mathbf{q} which are not small the calculation of $\epsilon(\mathbf{q}, \omega)$ is more complex, but introduces no qualitative change into the character of this quantity*.

* See, for example, D.Pines, *Elementary excitations in solids* (W.Benjamin, New York and Amsterdam, 1963).

Let us now see what the form of the field of a point charge is in a medium with dielectric constant (50.11). If the charge is at rest or moving slowly, then there corresponds to its potential in vacuum e/r the Fourier component

$$\varphi_0(\mathbf{q}) = 4\pi e^2/q^2. \quad (50.13)$$

The potential of a charge in a medium can be written in the form

$$\varphi(\mathbf{r}, t) = \int d\mathbf{q} \frac{\varphi(\mathbf{q})}{\epsilon(\mathbf{q}, 0)} = \frac{ee^{-r/l_D}}{r}, \quad (50.14)$$

where the Debye length l_D is equal to

$$l_D = \left(\frac{1}{4\pi e^2 \rho(\epsilon_F)} \right)^{\frac{1}{2}} = \left(\frac{\epsilon_F}{6\pi e^2 N} \right)^{\frac{1}{2}}. \quad (50.15)$$

Formula (50.14) shows that, as in a classical plasma, the field of a charge in a quantum plasma turns out to be screened. However, instead of the Debye radius $1/\kappa$, the screening length for a degenerate quantum plasma is given by formula (50.15).

Estimates show that l_D is about $1-2\text{\AA}$. Thus the screening effect of the plasma ensures the fall-off of the interaction forces at distances of the order of the mean distance between electrons in the metal. The screening of the electron charge has an obvious meaning: near any given electron the concentration of other electrons is reduced. This decrease in concentration is due to the pure Coulomb interaction as well as to exchange forces. The latter have figured implicitly in the calculation of the dielectric constant; that is, the use of the Fermi distribution corresponds to taking into account the Pauli principle and the exchange interactions associated with it.

A slow movement of an electron in a plasma is accompanied by a 'running away' of electrons from its path. Thus, not the electron but a whole group of particles, carrying a total charge equal to the charge of the electron, moves in the plasma. This system, consisting of the electron and a cloud of electrons moving together with it or away from it on a positive background formed by ions, can be considered as a quasiparticle with charge e and effective mass m^* . As will be seen from what follows, a contribution to the effective mass is given not only by the interaction between the electrons but also with the phonons of the lattice.

The existence of screening is of great importance for understanding the phenomena which occur in metals. Namely, owing to the screening, the

forces of interaction between electrons at relatively small distances are substantially reduced.

We shall return to the discussion of this problem somewhat later, and in the meanwhile we shall consider the form assumed by $\epsilon(\mathbf{q}, \omega)$ at high frequencies.

For this we transform (50.8) into a somewhat different form, writing for $\omega \rightarrow \infty$

$$\begin{aligned} & \frac{1}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} + \mathbf{q}) + \hbar\omega} + \frac{1}{\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} - \mathbf{q}) - \hbar\omega} = \\ & = \frac{2\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} + \mathbf{q}) - \epsilon(\mathbf{k} - \mathbf{q})}{[\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} + \mathbf{q}) + \hbar\omega][\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} - \mathbf{q}) - \hbar\omega]} \simeq \\ & \simeq - \frac{2\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} + \mathbf{q}) - \epsilon(\mathbf{k} - \mathbf{q})}{(\hbar\omega)^2}. \end{aligned}$$

Hence, if $\omega \rightarrow \infty$, we obtain for $\epsilon(\mathbf{q}, \omega)$

$$\epsilon(\mathbf{q}, \omega) \rightarrow 1 - \omega_p^2/\omega^2, \quad (50.16)$$

where ω_p denotes the quantity

$$\omega_p^2 = \frac{4\pi e^2}{\hbar^2 q^2} \int f(\mathbf{k}) [2\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} + \mathbf{q}) - \epsilon(\mathbf{k} - \mathbf{q})] d\mathbf{k}. \quad (50.17)$$

Expanding $\epsilon(\mathbf{k} + \mathbf{q})$ in a series, we can write

$$2\epsilon(\mathbf{k}) - \epsilon(\mathbf{k} + \mathbf{q}) - \epsilon(\mathbf{k} - \mathbf{q}) \simeq \frac{\partial^2 \epsilon}{\partial k^2} q^2 = \frac{\hbar^2}{m^*} q^2$$

so that

$$\omega_p^2 = 4\pi N e^2 / m^*.$$

Here we have made use of formula (49.14) for determining the effective mass. Comparing ω_p with the frequency of plasma oscillations of the classical plasma (46.16) of Part IV, we see that they are identical. The dispersion equation, defining the relation between the frequency and wave vector, according to (33.18) of Part IV is of the form $\epsilon(\mathbf{q}, \omega) = 0$. Here we assume the plasma

waves to be longitudinal (see §33 of Part IV). This gives $\omega = \omega_p$. Thus in a quantum plasma, as in a classical plasma, undamped (more precisely, undamped in our approximation) plasma oscillations can exist.

The meaning of this result is the same as in classical plasma theory: if a sufficiently large group of negatively charged particles of mass m^* is displaced with respect to a positively charged background, then the displaced particles will begin to oscillate about an equilibrium position with the plasma frequency.

If we did not confine ourselves to the first term of the expansion in the denominator of (50.8), then instead of (50.16) we would obtain a more complex dispersion equation from which it follows that plasma oscillations with frequencies different from ω_p are possible. The quanta of plasma oscillations are called plasmons.

Let us now see what the energy conditions for arousing plasma oscillations are. In order that an electron may give rise to collective oscillations in a system, i.e. emit a quantum with frequency ω_p and wave number q , it is necessary that the following energy conservation law be fulfilled:

$$\frac{\hbar^2 k^2}{2m} - \frac{\hbar^2 (\mathbf{k} - \mathbf{q})^2}{2m} = \hbar \omega_p$$

or

$$\frac{\hbar^2}{2m} (2\mathbf{k} \cdot \mathbf{q} - q^2) = \hbar \omega_p. \quad (50.18)$$

Since for electrons the wave vector $|\mathbf{k}|$ does not exceed its value at the Fermi surface, k_F , we see that condition (50.18) is not fulfilled for small values of q . This means that plasma oscillations in a system of free electrons can be brought about only for $q > \omega_p m / \hbar k_F = \omega_p / v_F$, where v_F is the velocity at the Fermi surface. The energy of a quantum of plasma oscillations $\hbar \omega_p \sim 20$ eV. This means that an individual electron with only thermal energy cannot give rise to plasma oscillations in a metal. This is not surprising, since plasma oscillations correspond to the collective motion of large groups of particles. If, however, a charged particle, for example an electron, possessing a sufficiently large energy, enters a metal, then such a particle may give rise to the plasma oscillations considered here.

This fact has made it possible to carry out successfully, experiments in which oscillations are excited in the plasma of solids. However, it should not be concluded from the above that plasma oscillations are absent in a system

of electrons in a metal. In contrast to the classical plasma, zero-point oscillations having energy $\frac{1}{2}\hbar\omega_p$ may exist in a quantum plasma. These are zero-point quanta or plasmons. Large groups of electrons take part in a correlated way in zero-point oscillations.

Let us estimate the order of magnitude of the size of such a region. The wavelength involving a group of particles, $\lambda_{\min} = 2\pi/q$, must be larger than the mean screening distance l_D . Using (50.18) and (50.15), we obtain $\lambda_{\min} > v_F/\omega_p$. No shorter plasma waves can be excited.

Thus it can be said that electrons at relatively large distances, larger than λ_{\min} , are involved in plasma waves.

We have now disposed of the essential information about a system of electrons in a solid, and are ready to make a more substantial evaluation of the role of the interactions between them in the general behaviour of the system. We write the part of the Hamiltonian characterizing the interaction between electrons in the form of two terms

$$H' = \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} = \sum_{\substack{i \neq j \\ |\mathbf{r}_i - \mathbf{r}_j| > \lambda_{\min}}} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{\substack{i \neq j \\ |\mathbf{r}_i - \mathbf{r}_j| < \lambda_{\min}}} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}.$$

The first sum involves only terms in which the distance between the particles is larger than λ_{\min} . At low electron energies $\epsilon < \epsilon_F$, this interaction gives rise to zero-point plasma oscillations. This sum brings a certain constant term into the total Hamiltonian. The second sum involves pairs inside the screening sphere. Such pairs interact with each other according to (50.14). Hence the corresponding term gives only a very small contribution to the total Hamiltonian. Calculation shows, for example, that the change in thermal capacity of the electron gas due to the screened interaction amounts to only a few percent. Thus the general conclusion which can be drawn from the above considerations is that, owing to the screening effect, the effect of the interaction between electrons on the properties of the solid turns out to be relatively small. This applies particularly to electrons having an energy close to the Fermi energy. These electrons always have the possibility of pushing the neighbouring electron cloud away from them and of moving together with this 'positive hole'.

Calculation shows that such a system has a considerable lifetime. This fact allows one to consider the system (electron + positive hole) to be a quasiparticle. The long lifetime is associated with the Pauli principle which prevents the electrons constituting the cloud from changing their state. The appearance of mobile quasiparticles is closely related to the thermal excitation of the sys-

tem, when the transitions of electrons into unoccupied states take place in it. Hence a system of electrons at $T \neq 0$ can be considered as an electron fluid in which elementary thermal excitations are moving. These excitations (quasi-particles), move independently of one another, have charge $(-e)$, mass m^* and spin $\frac{1}{2}$ and obey Fermi statistics. For brevity we shall call these quasi-particles the electrons.

We stress that disregarded factors, for instance the discrete distribution of positive charge, have no effect on the qualitative conclusions obtained.

The interaction of the electron (quasiparticle) with phonons, as we shall see below, leads to the formation around it of a cloud or 'jacket' of phonons, which moves together with it, changing its mass so that $m^* \rightarrow m^{**}$. The interaction between electrons is directly manifested in a number of effects, for example the difference in the velocity of sound in a metal in comparison with a dielectric. The character of this interaction shows why the model of an ideal gas of free electrons correctly gives the basic properties of a system of electrons in metals. The screening substantially reduces the interaction between electrons. The major effect of the interaction amounts to a change in the effective mass.

§51. Models of a metal, a semiconductor and a dielectric

We can now go on to a discussion of the properties of a system of electrons in a solid.

In a qualitative description of the behaviour of a system of electrons we can, on the basis of the results of the preceding section, replace it by a system of quasiparticles (fermions). In what follows we shall call these quasiparticles electrons, which, however, should not lead to misunderstanding. The number of non-localized electrons per cm^3 can be determined by direct measurements (see below). For the so-called 'good' metals such as the alkali and alkaline-earth elements, silver, copper, gold and a number of others, the number of non-localized electrons per cm^3 is approximately equal to the number of atoms per cm^3 , i.e. the number of such electrons per atom is equal to one.

For a number of other metals the number of 'free' electrons per atom turns out to be considerably less than one. Furthermore, a marked anisotropy is displayed, so that the properties of the system of electrons turn out to be different in different crystallographic directions. As an example, we mention bismuth, which exhibits a pronounced anisotropy in its electric and magnetic properties.

For metalloids the number of 'free' electrons per atom is so small that the

system of electrons forms a non-degenerate ideal gas. Finally, in a number of good dielectrics, for example in NaCl or solid oxygen, there are practically no free electrons. Although the density of free electrons is of decisive importance in the behaviour of the system of electrons in a solid, it would be incorrect to assume that the overall difference between metals, semiconductors and insulators amounts to the variation of this characteristic and is of a quantitative character. As a matter of fact, it has a profound meaning. This is seen, for example, from the qualitative difference in the mechanism of electric conduction or in the magnetic properties of metals, semiconductors and dielectrics. The profound difference in their physical properties is associated with the different character of the energy spectrum of their electrons.

We have seen that the electron spectrum in a crystal has the form of alternating bands of allowed and forbidden energies. It should be recalled that in each allowed energy band there is a limited number of states. Namely, there are $2N$ states in each band, where N is the number of unit cells per unit volume, and the factor 2 is associated with the two electrons occupying each state. If all the states of a band of allowed energies are occupied by electrons in pairs, then the electrons cannot change state and go from one state to another under the action of an applied external field. Such a body behaves as an insulator. But if in the allowed band only a fraction of the states are occupied by electrons, then transitions between states are possible and the body behaves as a metal.

Let us consider schematically the formation of a metal by separated non-interacting atoms. In these atoms let there be one valence electron in the highest occupied energy level. Its state will be two-fold degenerate, since the energy of an electron does not depend on the orientation of its spin. In a system of N independent non-interacting atoms the corresponding energy level will be $2N$ -fold degenerate. When the atoms are drawn together and an interaction is established between them, the level splits into $2N$ adjacent levels forming a continuous band. One half of these energy levels will be occupied by electron pairs, while the other half will be vacant. Thus this band of energy levels, called the conduction band, arises from the lower energy states of the valence electron. The excited states of the valence electron, splitting into a large number of levels, form other bands which are not occupied by electrons. In certain cases the broadening of bands is so large that they overlap each other. Between the allowed energy bands there are forbidden energy bands. The crystal formed will possess metallic properties, since the unoccupied states are directly adjacent to the occupied states. Such is the case for the alkali metals, copper, silver and some other metals.

Suppose now that the atom has two outer electrons in one energy state

with oppositely oriented spins. The states of the electrons in the atom will be non-degenerate. The corresponding state of a system of N non-interacting atoms will be $2N$ -fold degenerate. When the crystal is formed this state splits into $2N$ close levels occupied by electron pairs. A crystal with such an arrangement of levels is an insulator. In it the occupied energy levels are separated from the unoccupied ones by a region of forbidden energies with an interval $\Delta\epsilon$. In order that a thermal excitation may bring an electron from an occupied into an unoccupied state, the thermal energy kT must be of the order of ΔE . The same applies also to the excitation of electrons by an electric field. The values of the corresponding temperature or field strength turn out to be very large. For ordinary temperatures and fields, the electrons remain in occupied levels and cannot carry a current. Thus a dielectric differs from a metal not by the total number of electrons, but by the character of the arrangement of the allowed energy bands.

However, it should not be concluded from the above that atoms with two outer electrons form an insulator type crystal when they are combined.

In addition to the case discussed, there is possible a band structure in which, owing to the overlap of the bands arising from the normal and excited states of the atom, the unoccupied band is directly adjacent to the occupied one. Materials of this type are metals such as the alkaline-earth metals, lead and a number of others. Current theory does not enable one to predict which one of these two cases will result when atoms with the properties given are combined.

It goes without saying that the division of crystals into metals and dielectrics covers only two limiting cases. The whole range of intermediate properties between metals and dielectrics belongs to the semiconductors. In the case of semiconductors, the width of the forbidden zone is relatively small and becomes comparable with the energy of thermal excitation at a relatively low temperature. We shall dwell in more detail on the energy spectrum of semiconductors in §67.

§52. Magnetic properties of metals. The paramagnetism of an electron gas

It turns out that a number of important results can be obtained from the simplest model of a metal in which it is considered as a potential well with infinitely high walls, filled with a gas of free electrons. In particular, this crude scheme proves to be sufficient for discussing the magnetic properties of metals.

It turns out that the magnetic properties of metals are determined in the

first place by the behaviour of the non-localized electrons. The interaction of the electrons with the lattice has a relatively small effect on the magnetic properties of metals.

Let us consider the behaviour of a degenerate electron gas placed in a magnetic field. In such a system, two basic effects are shown. One of these is associated with the fact that electrons have spin, and the second with the quantization of the electron orbital motion in a magnetic field. We shall begin with the first effect.

When an external magnetic field is applied, preferential orientation of the spin moments takes place in the field. As a result of this, a magnetization arises in the system.

To calculate the magnetic susceptibility, we shall write, first of all, the expression for the free energy of the free electrons in a magnetic field. According to the Gibbs–Helmholtz formula (30.12) of Part III, we have

$$F = -T \int_0^T \frac{E dT}{T^2}, \quad (52.1)$$

where

$$E = \frac{1}{(2\pi\hbar)^3} \sum \int \frac{\epsilon d\gamma}{e^{(\epsilon - \mu + \mu_0 \cdot \mathbf{H})/kT} + 1}. \quad (52.2)$$

The summation is carried out over all (i.e. over two) possible orientations of the spin magnetic moment μ_0 with respect to the field \mathbf{H} . We recall that in the absence of a magnetic field the energy levels were degenerate and that the factor 2, instead of the summation, stood in the expression for the energy (80.2) of Part III.

Thus

$$E = \int \frac{\epsilon}{e^{(\epsilon - \mu + \mu_0 H)/kT} + 1} \frac{d\gamma}{(2\pi\hbar)^3} + \int \frac{\epsilon}{e^{(\epsilon - \mu - \mu_0 H)/kT} + 1} \frac{d\gamma}{(2\pi\hbar)^3}. \quad (52.3)$$

On substituting (52.3) into (52.1), we arrive at the calculation of integrals of

the type

$$\begin{aligned}
 \int \frac{\epsilon \, d\gamma}{(2\pi\hbar)^3} \int_0^T \frac{dT}{T^2(e^{(\epsilon-\mu+\mu_0 H)/kT} + 1)} &= \int \frac{d\gamma}{(2\pi\hbar)^3} k \int_{T^{-1}}^{\infty} \frac{\epsilon \, du}{e^{(\epsilon-\mu+\mu_0 H)u} + 1} = \\
 &= \int \frac{d\gamma}{(2\pi\hbar)^3} k \int_{T^{-1}}^{\infty} \frac{(\epsilon - \mu + \mu_0 H) \, du}{e^{(\epsilon-\mu+\mu_0 H)u} + 1} + \int \frac{d\gamma}{(2\pi\hbar)^3} k \int_{T^{-1}}^{\infty} \frac{(\mu - \mu_0 H) \, du}{e^{(\epsilon-\mu+\mu_0 H)u} + 1} = \\
 &= N(\mu - \mu_0 H) - k \int \frac{d\gamma}{(2\pi\hbar)^3} \ln(1 + e^{-(\epsilon-\mu+\mu_0 H)/kT}),
 \end{aligned}$$

where N is the number of particles. Hence for the free energy we can write

$$F = \mu N - kT \sum \int \frac{d\gamma}{(2\pi\hbar)^3} \ln(1 + e^{-(\epsilon-\mu+\mu_0 H)/kT}). \quad (52.4)$$

Let us consider the case of a weak magnetic field, for which the following inequality holds:

$$\mu_0 H/kT \ll 1. \quad (52.5)$$

Expanding (52.4) in a series in powers of this small quantity, we obtain

$$\begin{aligned}
 F &= N\mu - 2kT \int \frac{d\gamma}{(2\pi\hbar)^3} \ln(1 + e^{-(\epsilon-\mu)/kT}) + \\
 &+ \mu_0^2 H^2 kT \frac{\partial^2}{\partial \mu^2} \int \frac{d\gamma}{(2\pi\hbar)^3} \ln(1 + e^{-(\epsilon-\mu)/kT}).
 \end{aligned} \quad (52.6)$$

We calculate the integral

$$I = \int \frac{d\gamma}{(2\pi\hbar)^3} \ln(1 + e^{-(\epsilon-\mu)/kT}) = 2\pi \frac{(2m)^{\frac{3}{2}} V}{(2\pi\hbar)^3} \int_0^{\infty} \epsilon^{\frac{1}{2}} \ln(1 + e^{-(\epsilon-\mu)/kT}) \, d\epsilon.$$

Integrating by parts twice, we have

$$\begin{aligned} \int_0^{\infty} \epsilon^{\frac{1}{2}} \ln(1 + e^{-(\epsilon - \mu)/kT}) d\epsilon &= \epsilon^{\frac{1}{2}} \ln(1 + e^{-(\epsilon - \mu)/kT}) \Big|_0^{\infty} - \int_0^{\infty} \frac{\epsilon^{\frac{1}{2}} d\epsilon}{e^{(\epsilon - \mu)/kT} + 1} = \\ &= -\frac{2}{3kT} \int_0^{\infty} \epsilon^{\frac{3}{2}} \frac{\partial f}{\partial \epsilon} d\epsilon = \frac{4\mu^{\frac{5}{2}}}{15kT} \end{aligned}$$

by virtue of the property of the Fermi distribution (see (80.4) of Part III). Hence

$$I = \frac{8\pi(2m)^{\frac{3}{2}} V \mu^{\frac{5}{2}}}{15(2\pi\hbar)^3 kT}$$

and for the free energy we get

$$F = N\mu + \frac{16\pi(2m)^{\frac{3}{2}} V \mu^{\frac{5}{2}}}{15(2\pi\hbar)^3} - (\mu_0 H)^2 V \frac{8\pi(2m)^{\frac{3}{2}}}{15(2\pi\hbar)^3} \frac{\partial^2 \mu^{\frac{5}{2}}}{\partial \mu^2}. \quad (52.7)$$

According to (18.3) of Part IV, the magnetic susceptibility related to the orientation of spins is equal to

$$\chi_S = -\frac{1}{VH} \frac{\partial F}{\partial H} = \mu_0^2 \frac{4\pi(2m)^{\frac{3}{2}} \mu^{\frac{1}{2}}}{(2\pi\hbar)^3}. \quad (52.8)$$

The formula for the susceptibility can be written in another form:

$$\chi_S = \mu_0^2 n_{\text{eff}} / kT, \quad (52.9)$$

where for n_{eff} , the effective number of unpaired electrons, we have made use of formula (80.17) of Part III.

This last expression has an obvious meaning. Namely, (52.9) is the same as the susceptibility of a gas of n_{eff} particles which freely orient themselves in a magnetic field and have intrinsic magnetic moment μ_0 .

We see that the electron gas has a spin paramagnetism independent of temperature. It should be stressed that this result is closely associated with the degeneracy of the electron gas.

§ 53. The diamagnetism of an electron gas

It turns out, however, that in addition to spin paramagnetism an electron gas also displays orbital diamagnetism.

Orbital diamagnetism, discovered by L. Landau, does not have such an obvious origin as spin paramagnetism. To calculate the magnetic susceptibility associated with orbital motion, it is necessary to find the free energy of the electron gas in a magnetic field. For this it is, in turn, necessary to find the energy of a free electron in a uniform magnetic field.

Let us consider the solution of the Pauli equation in the simplest case of the motion of a free electron in a uniform magnetic field. We choose the direction of the field to be the z -axis, and write the vector potential in the form

$$A_x = -Hy; \quad A_y = A_z = 0.$$

For our purposes it is more convenient to write the vector potential in this form than in the form (19.16) of Part I. It is clear that the two expressions for the vector potential are equivalent.

The Pauli equation for stationary motion can be written as

$$\left\{ \frac{1}{2m} \left(p_x + \frac{e}{c} Hy \right)^2 + \frac{p_y^2}{2m} + \frac{p_z^2}{2m} - \frac{e}{mc} s_z H \right\} \psi = E \psi. \quad (53.1)$$

From eq. (53.1) it follows, first of all, that the coordinate wave function and spin wave function are independent.

An operator of the form $\text{const} \times s_z$ does not act on the variables x, y, z , so that eq. (53.1) can be satisfied by a wave function of the form

$$\psi = \varphi(s_z) \phi(x, y, z). \quad (53.2)$$

Since the coordinates x and z are not involved explicitly in eq. (53.1), we can try to find its solution in the form

$$\phi = e^{(i/\hbar)(p_x x + p_y z)} \zeta(y). \quad (53.3)$$

The momentum components p_x and p_y in the direction of the x -axis and z -

axis are conserved:

$$p_x H - H p_x = 0, \quad p_z H - H p_z = 0$$

and can run over a continuous sequence of values.

Substituting (53.3) into eq. (53.1), we obtain

$$\frac{\hbar^2}{2m} \xi''(y) + (E_1 - \frac{1}{2} m \omega^2 (y - y_0)^2) \xi(y) = 0, \quad (53.4)$$

where

$$\begin{aligned} E_1 &= E - \frac{p_z^2}{2m} + \mu_0 s_z H, \\ \omega &= eH/mc, \quad \mu_0 = eh/2mc, \\ y_0 &= -cp_x/eH. \end{aligned} \quad (53.5)$$

Eq. (53.4) is formally the same as the equation of motion of a linear harmonic oscillator with frequency ω (equal to the cyclotron frequency), oscillating about the equilibrium position y_0 . Hence, without reproducing the calculations of §10 of Part V, we can write

$$E_1 = (n + \frac{1}{2}) \hbar \omega \quad (53.6)$$

or

$$E = \frac{\hbar e H}{mc} (n + \frac{1}{2}) + \frac{p_z^2}{2m} - \mu_0 s_z H = 2\mu_0 H (n + \frac{1}{2}) + \frac{p_z^2}{2m} - \mu_0 s_z H \quad (53.7)$$

and

$$\xi(y) = H_n \{ \omega (y - y_0) \} e^{-(1/2) \omega^2 (y - y_0)^2}, \quad (53.8)$$

where H_n are Hermite polynomials.

Since the polynomials H_n decrease rapidly with increasing $\omega(y - y_0)$ and become small for $(y - y_0) > \sqrt{\hbar/m\omega}$ (see (10.2) and (10.15) of Part V), then, in a uniform magnetic field the particle moves freely in the direction

of the z -axis and performs a motion in the limited region

$$y_0 - (\hbar/m\omega)^{\frac{1}{2}} \leq y \leq y_0 + (\hbar/m\omega)^{\frac{1}{2}}. \quad (53.9)$$

The motion in this limited region corresponds to the classical motion of a charge in a circle with its cyclotron frequency. To the motion along the field there corresponds an energy $p_z^2/2m$, while to the motion in the xy -plane there corresponds the quantized energy E_{\perp} . The energy of the electron does not depend on the value of the momentum p_x , so that the states are degenerate.

We shall apply these results to the motion of an electron in the limited region (53.9) of space along the y -axis. In this case the momentum component p_x , in contrast to the component p_z , cannot take on arbitrary values. Indeed, the position of the 'equilibrium' point y_0 cannot be outside the limits of the size L of the region in the direction of the y -axis. Therefore, from (53.5) it follows that

$$0 \leq p_x \leq \frac{eH}{c} L. \quad (53.10)$$

Knowing the energy of an individual electron, which is composed of two independent parts: the energy of orbital motion, and the energy associated with the intrinsic magnetic moment, we can write the free energy of the electron gas. That is, we can set

$$F = F_0 + F_s + F_{\text{orb}}, \quad (53.11)$$

where F_s is the part of free energy due to the spin magnetic moment already found, and F_{orb} is the contribution of the orbital motion to the free energy

$$F_{\text{orb}} = -kT \int d\gamma' \sum_i \Omega_i \ln \left[1 + \exp \left\{ \left[\mu - 2\mu_0 H(n + \frac{1}{2}) - \frac{p_z^2}{2m} \right] / kT \right\} \right]. \quad (53.12)$$

The peculiarity of the system being considered lies in the fact that the momentum p_z in the direction of the field may vary continuously, whereas the motion in the xy -plane is quantized. Hence the integration is carried out with respect to the momentum p_z , and $d\gamma' = dz dp_z / 2\pi\hbar$. For given energy ϵ_n the states of orbital motion in the xy -plane are degenerate. The multiplicity of

degeneracy $\Omega(\epsilon_n)$ is equal to

$$\Omega(\epsilon_n) = 2 \frac{dx dy}{(2\pi\hbar)^2} 2\pi \int p_1 dp_1 = \frac{2\pi}{(2\pi\hbar)^2} \left(\int dx dy \right) p_1 \Big|_{p_1},$$

where the momentum p_1 corresponds to the motion in the xy -plane for a given energy E_1 .

In other words, it can be said that the multiplicity of degeneracy of the energy level E_1 is determined by a large but discrete (for finite size of the metal sample in the direction of the x -axis) number of possible values of p_x .

Writing the obvious relation

$$2\mu_0 H n < \frac{p_1^2}{2m} < 2\mu_0 H(n + \frac{1}{2})$$

we see that p_1 may vary (for given E_1) in the interval from $p_1 = (4m\mu_0 H n)^{\frac{1}{2}}$ to $p_2 = (4m\mu_0 H(n + \frac{1}{2}))^{\frac{1}{2}}$. Hence we find

$$\Omega(\epsilon_n) = \frac{8\pi m \mu_0 H}{(2\pi\hbar)^2} \int dx dy. \quad (53.13)$$

The same result follows from (53.10).

Substituting the values of $d\gamma'$ and $\Omega(\epsilon_n)$ into (52.12), we find

$$F_{\text{orb}} = -kT \frac{8\pi m \mu_0 H V}{(2\pi\hbar)^3} \times \\ \times \int_{-\infty}^{\infty} dp_z \sum_{n=0}^{\infty} \ln \left[1 + \exp \left\{ \left[\mu - 2\mu_0 H(n + \frac{1}{2}) - \frac{p_z^2}{2m} \right] / kT \right\} \right]. \quad (53.14)$$

Let us consider the case of weak fields, when inequality (52.5) is fulfilled. In this case the logarithmic function of the argument $\mu_0 H/kT$ changes slightly when n is replaced by $n + \frac{1}{2}$. Hence the summation can be replaced by integration using Euler's formula

$$\sum_{n=0}^{\infty} f(n + \frac{1}{2}) \simeq \int_0^{\infty} f(x) dx - \frac{1}{24} f'(x) \Big|_0^{\infty}.$$

By means of Euler's formula we get

$$\begin{aligned} \sum_{n=0}^{\infty} \ln \left[1 + \exp \left\{ \left[\mu - 2\mu_0 H(n + \tfrac{1}{2}) - \frac{p_z^2}{2m} \right] / kT \right\} \right] &\simeq \\ &\simeq \int_0^{\infty} dx \ln \left[1 + \exp \left\{ \left[\mu - 2\mu_0 Hx - \frac{p_z^2}{2m} \right] / kT \right\} \right] - \\ &- \frac{1}{24} \frac{2\mu_0 H}{kT} \left[\exp \left\{ \left(\frac{p_z^2}{2m} - \mu \right) / kT \right\} - 1 \right]^{-1}, \end{aligned}$$

Hence

$$\begin{aligned} F_{\text{orb}} &= -kT \frac{8\pi m \mu_0 H V}{(2\pi\hbar)^3} \times \\ &\times \int_{-\infty}^{\infty} dp_z \int_0^{\infty} dx \ln \left[1 + \exp \left\{ \left(\mu - 2\mu_0 Hx - \frac{p_z^2}{2m} \right) / kT \right\} \right] + \\ &+ \frac{2\mu_0 H \pi m \mu_0 H V}{3(2\pi\hbar)^3} \int_{-\infty}^{\infty} \left[\exp \left\{ \left(\frac{p_z^2}{2m} - \mu \right) / kT \right\} + 1 \right]^{-1} dp_z. \end{aligned} \quad (53.15)$$

In the first integral we introduce the new variable $u = 2\mu_0 Hx$, and in the second we set $\epsilon = p_z^2/2m$. Then we easily find

$$\begin{aligned} F_{\text{orb}} &= -\frac{kT(2\pi)(2m)^{\frac{3}{2}} V}{(2\pi\hbar)^3} \times \\ &\times \int_{-\infty}^{\infty} dp_z \int_0^{\infty} du \ln \left[1 + \exp \left\{ \left(\mu - u - \frac{p_z^2}{2m} \right) / kT \right\} \right] + \\ &+ \frac{2\pi(2m)^{\frac{3}{2}} (\mu_0 H)^2 V}{6(2\pi\hbar)^3} \int_0^{\infty} \frac{d\epsilon}{\epsilon^{\frac{1}{2}}} \frac{1}{e^{(\epsilon - \mu)/kT} + 1}. \end{aligned} \quad (53.16)$$

The first term of (53.16) does not depend on the magnetic field and gives no contribution to the magnetic susceptibility. According to (80.7) of Part III, the integral in the second term for a strongly degenerate electron gas can be written as

$$\int \frac{(\epsilon)^{-\frac{1}{2}} d\epsilon}{e^{(\epsilon-\mu)/kT}} \simeq 2\mu^{\frac{1}{2}}.$$

The susceptibility of an electron gas associated with the orbital motion turns out to be equal to

$$\chi_{\text{orb}} = -\frac{1}{VH} \frac{\partial F_{\text{orb}}}{\partial H} = -\frac{4\pi(2m)^{\frac{3}{2}} \mu_0^2 \mu^{\frac{1}{2}}}{3(2\pi\hbar)^3}. \quad (53.17)$$

We see that the orbital magnetic susceptibility is negative, i.e. corresponds to the diamagnetism of the electron gas.

Comparing (53.17) and (52.8), it is easily seen that

$$|\chi_{\text{orb}}| = \frac{1}{3}\chi_s.$$

Hence the total susceptibility of an electron gas turns out to be equal to

$$\chi = \frac{2}{3}\chi_s = \frac{8\pi(2m)^{\frac{3}{2}} \mu_0^2 \mu^{\frac{1}{2}}}{3(2\pi\hbar)^3}. \quad (53.18)$$

This value of χ proves to be in good agreement with experimental data.

A more detailed treatment taking into account the effect of the crystal lattice does not affect the general result. We stress that in calculating the diamagnetic susceptibility we assumed the magnetic field to be weak. In strong magnetic fields, when inequality (52.5) is not fulfilled, the magnetic susceptibility turns out to be a function of the field strength. This function has an oscillatory character (de Haas—van Alphen effect). The oscillations are associated with the change of energy E_1 with the field and the corresponding change in occupations of the Fermi levels*.

* See, for example, J.M.Ziman, *Principles of the theory of solids* (Cambridge University Press, Cambridge, 1964).

§54. Ferromagnetism

As was stressed in Part IV, ferromagnetic properties are possessed by only a relatively small number of metals. That is, only some metals with unfilled inner shells (transition metals) and some alloys turn out to be ferromagnetic.

Among the macroscopic properties of ferromagnetic bodies the existence of a very large permanent magnetization is characteristic.

The main feature of a more detailed description of ferromagnetic bodies is the correlation of the orientations of magnetic moments.

From this point of view, such properties are possessed by antiferromagnetic and ferromagnetic substances as well as by ferromagnetics. In other words, in the broad sense all strongly magnetized substances are ferromagnetic. Measurements of the ratio of the magnetic moment to angular momentum have shown that for ferromagnetic substances it is equal to e/mc , i.e. that it corresponds to the spin nature of the magnetic moment. This suggests that strong magnetization is associated with the interaction of the spin moments of the electrons.

The set of two basic facts; the spin nature of magnetization, and the decisive role of open atomic shells, leads to the following picture of the phenomenon: the energy spectrum of electrons in ferromagnetic metals has two bands. One of these, formed by strongly interacting valence electrons, has a large width and is a conduction band. The other arises from the interaction of electrons in unfilled shells. This interaction is relatively weak, since the overlapping of the wave functions of the inner electrons is small. In correspondence with formula (49.13), the width of this band is small and it cannot be responsible for conduction.

However, we have already seen (§19 of Part IV) that the spin of electrons of unfilled shells is easily reversed. Hence a relatively weak interaction between the electrons may lead to alignment of their spins.

Detailed consideration of this interaction is a complex problem. We shall confine ourselves to the discussion of the following model of a ferromagnetic material. Let there be a lattice of regularly arranged atoms, having one electron in an s-state whose spin can have two orientations. In the absence of interaction the symmetrized wave function of a system of N electrons has the form (see (65.6) of Part V)

$$\Psi = (N!)^{-1/2} \sum (-1)^n P_n \left[\prod_j \psi_j(\mathbf{r}_j) \varphi_j(\uparrow) \right], \quad (54.1)$$

where the vector index \mathbf{j} denotes the number of the atom, and $\varphi(\uparrow)$ denotes the spin wave function corresponding to the 'regular' spin orientation.

Suppose that in the ground state of the system all the spins have one and the same orientation, for example $+\frac{1}{2}$. Such a completely ordered orientation of spins corresponds to the maximum magnetization of the system. We shall see below under what condition such a state of a system corresponds to its minimum energy.

The interaction between electrons leads to a disturbance of the regularity of the orientation of their spins. Some of the spins turn out to be oppositely oriented. The perturbation being small means that the number of oppositely oriented spins is very small in comparison with the total number of spins.

Let us consider a perturbed state in which the spin of one electron, for instance of the m th atom, is reversed.

The wave function of such a state can be written in the form

$$\Psi_m = (N!)^{-\frac{1}{2}} \sum (-1)^n P_n \psi_m(r_m) \varphi_m(\downarrow) \prod_j' \psi_j(r_j) \varphi_j(\uparrow), \quad (54.2)$$

where the prime on the product sign denotes the absence of one factor. Since the system is degenerate (i.e. the reversed spin can belong to any atom) then according to the general rules of perturbation theory for degenerate systems the wave function must be written in the form

$$\Psi = \sum C_m \Psi_m. \quad (54.3)$$

The subsequent calculation is the same as that given in §49. The overall difference amounts to taking into account the antisymmetry of the wave function (54.2). Since the overlap of the wave functions is small, we shall restrict ourselves to taking into account the interaction of electrons of only the nearest neighbouring atoms (i.e. to the interaction of the electron of the m th atom with the electrons of the $(m-1)$ th and $(m+1)$ th atoms).

Then, analogously to (49.8), we can write the system of linear equations

$$(E' - E_0 + \alpha)C_m + \sum_{m'} \beta(m') C_{m+m'} = 0, \quad (54.4)$$

where E' is the energy of the perturbed state, E_0 is the energy of the ground state, and α and β are the Coulomb and exchange integrals,

$$\alpha = - \int \psi_m^* U' \psi_m dV,$$

$$\beta = \beta_{m,m+1} = \beta_{m,m-1} =$$

$$= - \int \psi_{m+1}^*(\mathbf{r}_m) \psi_m^*(\mathbf{r}_{m+1}) U' \psi_{m+1}(\mathbf{r}_m) \psi_m(\mathbf{r}_{m+1}) dV_m dV_{m+1}.$$

Here U' is the energy of interaction between the electron of the m th atom and the electrons (atom core) of the $(m+1)$ th or $(m-1)$ th atom.

The solution of (54.4) reads (see (49.9))

$$c_m = e^{i\mathbf{k} \cdot \mathbf{m}}, \quad (54.5)$$

where the vector \mathbf{k} is defined by the conditions of periodicity. Hence the wave function is

$$\Psi = \sum e^{i\mathbf{k} \cdot \mathbf{m}} \psi_m. \quad (54.6)$$

For the excited state energy we obtain in the case of a simple cubic crystal

$$E' = E_0 - \alpha - 2\beta(\cos k_x a + \cos k_y a + \cos k_z a) \quad (54.7)$$

where a is the lattice spacing.

To small excitation energies there correspond small values of the vector \mathbf{k} , so that

$$E' = E_0 - \alpha - 6\beta - \frac{1}{2}\beta a^2 k^2 = E_0 - \alpha - 6\beta - \frac{\hbar^2 k^2}{2m^*}; \quad m^* = \hbar/\beta a^2. \quad (54.8)$$

The wave function (54.6) and energy (54.8) permit an obvious interpretation.

We see that the excitation energy is connected with the exchange interaction of electrons. For the excited state energy to be higher than the ground state energy it is necessary for the following condition to be fulfilled, $\beta > 0$, i.e. the exchange integral must be positive. There corresponds to the excited state of the crystal the presence of a deviation of the spin orientation. Since all atoms of the crystal are equivalent, the spin deviation is not fixed to a definite atom but may wander throughout the crystal.

The displacement of the spin deviation in the crystal from one atom to another is described by the wave function (54.6); the energy spectrum of the crystal is given by formula (54.7). The wave function (54.6) is called the spin wave. Formally the displacement of the spin deviation can be compared to the motion of a certain quasiparticle, called a magnon. Indeed, formally, the spin wave describes the propagation of a certain quasiparticle with wave vector \mathbf{k} in the crystal. The energy of this quasiparticle is given by formula

(54.8), where m^* is its effective mass. Magnons represent elementary excitations in a system of oriented spins. The energy of excitation of the entire crystal can be considered as the sum of such elementary excitations or as the energy of an ideal gas of magnons filling the entire volume of the crystal.

It is clear that such a description of the excited state is approximate and is reasonable only for small degrees of excitation. This means that the number of magnons (the average number of spins excited into a reversed state) must be sufficiently small. Somewhat later we shall give quantitative character to this statement.

All that we said in §47 apropos the description of the excited state of a system of particles by means of quasiparticles also applies to magnons. Magnons can interact with other quasiparticles, for example with phonons; or with real particles possessing magnetic moments, for example with neutrons. Magnons must be considered to obey Bose–Einstein statistics. Indeed, any number of electron spins belonging to different atoms may be in a state with given orientation. The number of magnons, like the number of phonons, is not conserved. To the excitation of the spin of one or another electron into a ‘reversed’ state there corresponds the appearance (or disappearance) of a magnon.

Hence the distribution function for the number of magnons with a given wave vector is represented by a Planck distribution

$$f(\mathbf{k}) d\mathbf{k} = \frac{1}{(2\pi)^3} \frac{d\mathbf{k}}{e^{\hbar^2 k^2 / 2m^* kT} - 1}. \quad (54.9)$$

Knowing that the character of the excitations in a crystal is associated with the exchange interaction of electrons with a free spin, we can come back to the discussion of the magnetic properties of such crystals.

If the exchange integral is positive, then in the ground state all spins are oriented in one direction. In this case the crystal possesses the spontaneous saturation magnetic moment

$$M_0 = \mu_0 N,$$

where N is the number of electrons per unit volume.

At $T \neq 0$ some of the spins will be reversed. Clearly, the number of reversed spins is equal to the number of magnons excited. The latter, according to (54.9), is equal to

$$N_{\text{magn}} = \frac{4\pi}{(2\pi\hbar)^3} \int \frac{k^2 dk}{e^{\hbar^2 k^2 / 2m^* kT} - 1}.$$

For low temperatures the integral converges rapidly, and the range of integration can be replaced by an infinite range. This gives

$$\begin{aligned}
 N_{\text{magn}} &= \frac{4\pi}{(2\pi)^3} \int_0^\infty \frac{k^2 dk}{e^{\hbar^2 \pi^2 / 2m^* kT} - 1} = \\
 &= \frac{1}{2\pi^2} \left(\frac{2m^* kT}{\hbar^2} \right)^{\frac{3}{2}} \int_0^\infty \frac{x^2 dx}{e^{x^2} - 1} \simeq \frac{1.3}{\pi^2 a^3} \left(\frac{kT}{\beta} \right)^{\frac{3}{2}} \simeq \frac{1.3}{\pi^2} N \left(\frac{kT}{\beta} \right)^{\frac{3}{2}}.
 \end{aligned} \tag{54.10}$$

Here we have made use of an approximate value of the integral and replaced $1/a^3$ by the number of electrons per unit volume, and substituted for m^* its value from (54.8).

The excitation of N_{magn} magnons reduces the magnetic moment of the crystal. It turns out to be equal to

$$M = M_0 \left[1 - \frac{1.3}{\pi^2} \left(\frac{kT}{\beta} \right)^{\frac{3}{2}} \right]. \tag{54.11}$$

Formula (54.11) defines the spontaneous magnetic moment of a ferromagnet as a function of temperature. At a certain temperature T_C , called the Curie point, the spontaneous magnetic moment must reduce to zero. It is clear that

$$T_C \approx \pi\beta/k.$$

If the exchange integral β is assumed to be of the order of magnitude of e^2/a , where a is the lattice constant, then $T_C \sim 10^3$ K, which is in agreement with experimental data. The temperature dependence of the spontaneous magnetic moment given by formula (54.11) is also in good agreement with experimental data for low temperatures ($T \ll T_C$).

For temperatures $T \sim T_C$ this theory loses any quantitative meaning, since the number of magnons becomes of the same order of magnitude as the total number of electrons.

In spite of its extremely sketchy character, the theory of ferromagnetism given above not only explains the essence of the phenomenon but also allows one to draw certain quantitative conclusions.

We have left out of account the anisotropy of the system associated with the anisotropy of the wave function of electrons in the unfilled d-shell. Also, we have not taken into consideration the spin-spin and spin-orbit interac-

tions and the interaction with the electric field of the lattice, which also give rise to anisotropy and lead to the appearance of directions of easy magnetization. These effects are experimentally rather well investigated, but their theory, which is associated with great mathematical difficulties, is to a great extent still not worked out. Referring the reader for details to more specialized texts†, we shall deal with only one problem here, that of the sign of the exchange integral.

As we have seen above, the condition $\beta > 0$ is of decisive importance for the existence of ferromagnetism. The exchange integral β can be written in the form

$$\beta = e^2 \int \psi_{n_1}^*(1) \psi_{n_2}^*(2) \left[\frac{1}{r_{12}} - \frac{1}{r_{1,n_1}} - \frac{1}{r_{2,n_2}} \right] \psi_{n_1}(2) \psi_{n_1}(1) dV_1 dV_2.$$

The first term in the bracket gives the Coulomb interaction between the electrons, and the two remaining terms give the interaction of the electrons with the nuclei. For the integral to be positive, the intrinsic exchange energy must be large. This means that the wave function must be large at a large distance from the nucleus and small close to it. This condition is satisfied by electrons in d-states, which have a relatively large angular momentum.

At the same time, the radius of the orbits of those electrons which are responsible for the interactions must be small in comparison with the lattice constant. Otherwise the electrons will closely approach 'strange' nuclei and the negative terms will give a large contribution to the integral.

These conditions exist for the atoms of the iron group, which are the most characteristic representatives of ferromagnetic substances. However, these rather stringent conditions are more often not fulfilled and the integral turns out to be negative. In this case the exchange interaction of the electrons of open shells leads to another kind of ground state. That is, in the ground state, neighbouring spins turn out to be antiparallel and spontaneous magnetization is absent.

Such substances are called antiferromagnetic. Antiferromagnetics, possessing no spontaneous magnetization, display a number of characteristic properties. We cannot, however, dwell on a consideration of these here, and refer the reader to more specialized texts††.

† S.V.Tyablikov, *Metody kvantovoi teorii ferromagnetizma (Methods of the quantum theory of ferromagnetism)* (Nauka, Moscow, 1965); J.M.Ziman, *Principles of the theory of solids* (Cambridge University Press, Cambridge, 1964).

†† J.M.Ziman, *Principles of the theory of solids* (Cambridge University Press, Cambridge, 1964).

§55. The interaction of electrons with lattice vibrations

We have discussed above the problem of the interaction of an electron with a crystal for a regular arrangement of ions at lattice points. The motion in the lattice of an electron and the hole associated with a quasiparticle does not differ from the motion of a single electron. The quasiparticle in the lattice is described by the same Bloch function.

In what follows we shall discuss the problem of the interaction of the quasiparticle with lattice vibrations. In accordance with what was said at the end of §50, we shall call the quasiparticle the electron without further specifying it.

Thermal lattice vibrations violate the periodicity of the potential acting on the electron. As the lattice vibrates the atoms undergo displacements from their equilibrium positions, which we characterize by the vector ξ_n .

We are interested in the value of the potential energy U of the electron at a certain point \mathbf{r} of the lattice. It can be found for two limiting cases: in the approximation of deformable ions and in the approximation of rigid, undeformable ions. The two models lead to very similar results. We shall make use of the first model which appears to be somewhat closer to the true state of affairs. In the model of deformable ions the entire lattice is replaced by a vibrating continuum. At a given instant of time the potential energy of the electron at the point \mathbf{r} is determined by the instantaneous configuration of this continuum. To the displacements of the continuum there will correspond changes in the potential energy of the electron at the point \mathbf{r} . That point of the continuum which in the equilibrium position had the coordinate \mathbf{r} , after the displacement will go over into point $\mathbf{r} + \xi(\mathbf{r})$. The value of the potential energy of the electron which in equilibrium corresponded to the radius vector $\mathbf{r} - \xi(\mathbf{r})$, after the displacement will correspond to the radius vector \mathbf{r} . In other words, that potential energy which was originally 'located' at the point $\mathbf{r} - \xi(\mathbf{r})$ will 'move' to the point \mathbf{r} . Obviously, the change in the potential energy of the electron at the point \mathbf{r} will be

$$U_{\text{latt}} = U(\mathbf{r} - \xi) - U(\mathbf{r}) = -\xi(\mathbf{r}) \nabla U, \quad (55.1)$$

if the displacement is assumed to be sufficiently small. It is natural to assume that for the vector $\xi(\mathbf{r})$ one can write an expression in which the vector \mathbf{n} is replaced by the vector \mathbf{r} .

Let us consider by means of the method of second quantization a system of electrons each of which is acted upon by the perturbation U_{latt} . We shall disregard the Coulomb interaction of the electrons with each another. In cor-

respondence with the results given in §99 of Part V, we can write the Hamiltonian of the system of electrons in the external lattice field. Namely, we shall characterize the state of the system of electrons by the occupation numbers of states with a given value of \mathbf{k} .

Using the second quantization representation (formula (99.23) of Part V) we write the energy of the system of electrons in the form

$$\hat{H} = \sum_{\mathbf{k}} E_{\mathbf{k}} \hat{a}_{\mathbf{k}}^{\dagger} \hat{a}_{\mathbf{k}} + \sum_{\mathbf{k}, \mathbf{k}'} \hat{a}_{\mathbf{k}}^{\dagger} \hat{a}_{\mathbf{k}'} (\mathbf{k}' | U_{\text{latt}} | \mathbf{k}) = \hat{H}_0 + \hat{H}_{\text{int}}.$$

Here $E_{\mathbf{k}}$ represents the energy of an electron moving in a strictly periodic lattice field. The operators $\hat{a}_{\mathbf{k}}^{\dagger}$ and $\hat{a}_{\mathbf{k}}$ satisfy the usual anticommutation relations (99.29) of Part V.

The interaction operator is determined by the matrix element of the energy of interaction of the electrons with an external field, in the given case, the field of a vibrating lattice. According to formula (99.23) of Part V

$$(\mathbf{k}' | U_{\text{latt}} | \mathbf{k}) = \int \psi_{\mathbf{k}'}^* U_{\text{latt}} \psi_{\mathbf{k}} dV.$$

Hence the Hamiltonian is of the form

$$\hat{H} = \sum_{\mathbf{k}} E_{\mathbf{k}} \hat{a}_{\mathbf{k}}^{\dagger} \hat{a}_{\mathbf{k}} + \sum_{\mathbf{k}'} \sum_{\mathbf{k}} \hat{a}_{\mathbf{k}}^{\dagger} \hat{a}_{\mathbf{k}'} \int \psi_{\mathbf{k}'}^* U_{\text{latt}} \psi_{\mathbf{k}} dV. \quad (55.2)$$

Substituting into (55.2) the wave functions $\psi_{\mathbf{k}}$ defined by formula (48.11) and the operator U_{latt} from (55.1), we obtain for the interaction operator

$$\hat{H}_{\text{int}} = -\frac{1}{N} \sum_{\mathbf{k}} \sum_{\mathbf{k}'} \hat{a}_{\mathbf{k}}^{\dagger} \hat{a}_{\mathbf{k}'} \int u_{\mathbf{k}}^* e^{-i\mathbf{k} \cdot \mathbf{r}} (\hat{\xi} \nabla U) u_{\mathbf{k}'} e^{i\mathbf{k}' \cdot \mathbf{r}} dV.$$

Making use of the explicit form of the operator $\hat{\xi}_n$ (47.14), we write the interaction operator in the form

$$\begin{aligned} \hat{H}_{\text{int}} = & -\frac{1}{N^{\frac{3}{2}}} \sum_{\mathbf{k}'} \sum_{\mathbf{k}} \sum_{j=1}^3 \hat{a}_{\mathbf{k}}^{\dagger} \hat{a}_{\mathbf{k}'} \int u_{\mathbf{k}}^* e^{-i\mathbf{k} \cdot \mathbf{r}} \\ & \times \sum_{fj} \left(\frac{\hbar}{2M\omega_f} \right)^{\frac{1}{2}} e_{fj} (\hat{b}_{fj} e^{i\mathbf{f} \cdot \mathbf{r}} + \hat{b}_{fj}^{\dagger} e^{-i\mathbf{f} \cdot \mathbf{r}}) \nabla U e^{i\mathbf{k} \cdot \mathbf{r}} u_{\mathbf{k}'} dV. \end{aligned} \quad (55.3)$$

We now pass from integration over the basic region to integration over a

unit cell of the crystal. By means of relation (49.18) we obtain for the two integrals involved in (55.3)

$$\int_G u_{\mathbf{k}}^* \nabla U_{\mathbf{k}} e^{i(\mathbf{k}-\mathbf{k}' \pm \mathbf{f}) \cdot \mathbf{r}} dV = \sum_n e^{i(\mathbf{k}-\mathbf{k}' \pm \mathbf{f}) \cdot \mathbf{n}} \int_{\tau_0} u_{\mathbf{k}}^* \nabla U u_{\mathbf{k}} dV. \quad (55.4)$$

Making use of formula (49.19), we see that integral (55.4) differs from zero only if the equality

$$\mathbf{k} - \mathbf{k}' \pm \mathbf{f} = 0 \quad (55.5)$$

or, in the more general case, the equality

$$\mathbf{k} - \mathbf{k}' \pm \mathbf{f} = \mathbf{K}, \quad (55.6)$$

is fulfilled. We shall confine ourselves to the consideration of the first case. Formula (55.5) shows that in the process of the scattering of an electron by a vibrating lattice a conservation law holds analogous to the momentum conservation law in the collision of free particles. Hence the process of scattering can formally be considered as the process of absorption or emission of a phonon by the electron. Before the collision the electron had wave number \mathbf{k} ; after the collision it has wave number $\mathbf{k}' = \mathbf{k} \pm \mathbf{f}$, where \mathbf{f} is the wave vector of the phonon. It can be said that the absorption (for $\mathbf{k}' = \mathbf{k} + \mathbf{f}$) or emission (for $\mathbf{k}' = \mathbf{k} - \mathbf{f}$) of the phonon occurs in the transition $\mathbf{k} \rightarrow \mathbf{k}'$.

We shall see below that in this case the energy conservation law holds, so that in the scattering the energy of the electron increases or decreases by the value of the phonon energy $\hbar\omega_{\mathbf{f}}$

$$E_{\mathbf{k}'} = E_{\mathbf{k}} \pm \hbar\omega_{\mathbf{f}}. \quad (55.7)$$

This obvious treatment of the process of the scattering of electrons by lattice thermal vibrations turned out to be very useful.

Collisions for which equality (55.6) holds are called Umklapp processes. It is easily seen that for small values of \mathbf{f} and values of the electron wave vector $|\mathbf{k}|, |\mathbf{k}'| \leq \pi/a$ the Umklapp process corresponds to a change of direction of the electron vector into the opposite direction.

The Umklapp processes do not play any special role in the phenomena of electric conduction considered below but are important for establishing thermal equilibrium in metals, particularly at low temperatures.

By means of relation (49.19), we obtain for (55.4)

$$\int_G u_{\mathbf{k}}^* e^{i(\mathbf{k}-\mathbf{k}) \cdot \mathbf{r}} (\nabla U) u_{\mathbf{k}} dV = N \int_{\tau_0} u_{\mathbf{k} \pm \mathbf{f}}^* (\nabla U) u_{\mathbf{k}} dV.$$

Substituting this value into (55.3), we find

$$\begin{aligned} \hat{H}_{\text{int}} &= - \sum_{\mathbf{k}, \mathbf{f}, j} \left(\frac{\hbar}{2MN\omega_f} \right)^{\frac{1}{2}} \mathbf{e}_{fj} \cdot \left\{ \left[\int_{\tau_0} u_{\mathbf{k}+\mathbf{f}}^* (\nabla U) u_{\mathbf{k}} dV \right] \hat{a}_{\mathbf{k}+\mathbf{f}}^\dagger \hat{a}_{\mathbf{k}} \hat{b}_{fj} + \right. \\ &\quad \left. + \left[\int_{\tau_0} u_{\mathbf{k}-\mathbf{f}}^* (\nabla U) u_{\mathbf{k}} dV \right] \hat{a}_{\mathbf{k}-\mathbf{f}}^\dagger \hat{a}_{\mathbf{k}} \hat{b}_{fj} \right\} = \\ &= - \sum_{\mathbf{k}, \mathbf{f}, j} \left(\frac{\hbar}{2MN\omega_f} \right)^{\frac{1}{2}} (\hat{a}_{\mathbf{k}+\mathbf{f}}^\dagger \hat{a}_{\mathbf{k}} \hat{b}_{fj} S_+ + \hat{a}_{\mathbf{k}-\mathbf{f}}^\dagger \hat{a}_{\mathbf{k}} \hat{b}_{fj} S_-), \end{aligned} \quad (55.8)$$

where S_{\pm} denotes integrals over a unit cell

$$S_{\pm} = \mathbf{e}_{fj} \cdot \int_{\tau_0} u_{\mathbf{k} \pm \mathbf{f}}^* \nabla U u_{\mathbf{k}} dV. \quad (55.8')$$

Formula (55.8) again allows an obvious treatment of the scattering of the electron as a process in which the absorption or emission of a phonon occurs. Indeed, from the meaning of the operators \hat{a}^\dagger , \hat{a} , \hat{b}^\dagger and \hat{b} it is clear that the matrix element of the operator $\hat{a}_{\mathbf{k}+\mathbf{f}}^\dagger \hat{a}_{\mathbf{k}} \hat{b}_{fj}$ corresponds to the appearance of an electron with wave number $\mathbf{k} + \mathbf{f}$ and to the disappearance of an electron with wave number \mathbf{k} and of a phonon with wave number \mathbf{f} . Analogously, the matrix element $\hat{a}_{\mathbf{k}-\mathbf{f}}^\dagger \hat{a}_{\mathbf{k}} \hat{b}_{fj}^\dagger$ corresponds to the appearance of an electron with wave number $\mathbf{k} - \mathbf{f}$ and of a phonon with wave number \mathbf{f} and to the disappearance of an electron with wave number \mathbf{k} . Using the property of periodicity of the potential U and of the functions $u_{\mathbf{k}}$, $u_{\mathbf{k}+\mathbf{f}}$, as well as the Schrödinger equation for $u_{\mathbf{k}}$, then on the basis of certain simplifying assumptions the unknown potential energy U can be eliminated from S_{\pm} . Calculations which we shall carry out below give for S_{\pm}

$$S_{\pm} = \pm \frac{i\hbar^2}{3m} (\mathbf{f} \cdot \mathbf{e}_{fj}) \int_{\tau_0} |\nabla u_{\mathbf{k}}|^2 dV. \quad (55.9)$$

It follows from (55.9) that the quantity S_{\pm} differs from zero only for lat-

tice waves with longitudinal polarization. Indeed,

$$\mathbf{e}_{f1} \cdot \mathbf{f} = f, \quad \mathbf{e}_{f2} \cdot \mathbf{f} = \mathbf{e}_{f3} \cdot \mathbf{f} = 0,$$

where index 1 refers to longitudinal polarization, and indices 2 and 3 refer to transverse polarization. Thus, within the framework of the assumptions made in calculating S (see below), the scattering of electrons takes place only on lattice waves with longitudinal polarization.

Substituting the value of S_z from (55.9) into the interaction energy operator (55.8), one can write it in the form

$$\begin{aligned} \hat{H}_{\text{int}} = & -i \sum_{\mathbf{k}, \mathbf{f}} \frac{\hbar^2 f}{3m} \left(\frac{\hbar}{2MN\omega_f} \right)^{\frac{1}{2}} \left(\int_{\tau_0} |\nabla u_{\mathbf{k}}|^2 dV \right) \times \\ & \times (\hat{a}_{\mathbf{k}+\mathbf{f}}^\dagger \hat{a}_{\mathbf{k}} \hat{b}_{\mathbf{f}} - \hat{a}_{\mathbf{k}-\mathbf{f}}^\dagger \hat{a}_{\mathbf{k}} \hat{b}_{\mathbf{f}}^\dagger). \end{aligned} \quad (55.10)$$

For brevity we omit the index $j=1$ on the operators \hat{b} . A further simplification is obtained if ω_f and \mathbf{f} are related by formula (47.18) which is valid for small \mathbf{f} , i.e. if frequency dispersion is disregarded. Then we obtain for \hat{H}_{int}

$$\hat{H}_{\text{int}} = -i \sum_{\mathbf{k}, \mathbf{f}} D_\omega (\hat{a}_{\mathbf{k}+\mathbf{f}}^\dagger \hat{a}_{\mathbf{k}} \hat{b}_{\mathbf{f}} - \hat{a}_{\mathbf{k}-\mathbf{f}}^\dagger \hat{a}_{\mathbf{k}} \hat{b}_{\mathbf{f}}^\dagger), \quad (55.11)$$

where

$$D_\omega = \left(\frac{2g^2 \hbar \omega_f}{9MNc^2} \right)^{\frac{1}{2}}; \quad g = \frac{\hbar^2}{2m} \int_{\tau_0} |\nabla u_{\mathbf{k}}|^2 dV. \quad (55.12)$$

In the next section we shall make use of expression (55.11) to calculate the transition probability.

We note that in order of magnitude g represents the mean kinetic energy of an electron moving in the lattice.

In conclusion we have to prove formula (55.9). For this we consider the integral

$$\int_{\tau_0} \nabla \cdot (\mathbf{e}_{fj} u_{\mathbf{k}}^* u_{\mathbf{k}} U) dV = \int_S \mathbf{e}_{fj} u_{\mathbf{k}}^* u_{\mathbf{k}} U dS = 0,$$

which reduces to zero by virtue of the properties of periodicity of the func-

tions $u_{\mathbf{k}}$ and of the potential U at the surface. On the other hand, this integral is equal to

$$\begin{aligned} \int_{\tau_0} \nabla \cdot (\mathbf{e}_{fj} u_{\mathbf{k}}^*, u_{\mathbf{k}} U) dV = \\ = \mathbf{e}_{fj} \cdot \int U [\nabla(u_{\mathbf{k}}^* u_{\mathbf{k}})] dV + \mathbf{e}_{fj} \cdot \int u_{\mathbf{k}}^* u_{\mathbf{k}} (\nabla U) dV = 0. \end{aligned} \quad (55.13)$$

Thus for quantity S (55.8') we obtain

$$S = \mathbf{e}_{fj} \cdot \int u_{\mathbf{k}}^* u_{\mathbf{k}} (\nabla U) dV = - \mathbf{e}_{fj} \cdot \int U [\nabla(u_{\mathbf{k}}^* u_{\mathbf{k}})] dV. \quad (55.14)$$

The potential energy U can now be eliminated by means of the Schrödinger equation. Using formula (48.9) for the function $u_{\mathbf{k}}$, we have

$$\frac{\hbar^2}{2m} [\nabla^2 u_{\mathbf{k}} + 2i\mathbf{k} \cdot \nabla u_{\mathbf{k}} - (k)^2 u_{\mathbf{k}}] + [E(\mathbf{k}) - U] u_{\mathbf{k}} = 0. \quad (55.15)$$

For the function $u_{\mathbf{k}}^*$, we write the analogous equation

$$\frac{\hbar^2}{2m} [\nabla^2 u_{\mathbf{k}}^* + 2i\mathbf{k}' \cdot \nabla u_{\mathbf{k}}^* - (k')^2 u_{\mathbf{k}}^*] + [E(\mathbf{k}') - U] u_{\mathbf{k}}^* = 0. \quad (55.16)$$

We multiply eq. (55.15) by the quantity $\mathbf{e}_{fj} \cdot \nabla u_{\mathbf{k}}^*$, and eq. (55.16) by $\mathbf{e}_{fj} \cdot \nabla u_{\mathbf{k}}$, and add the equations obtained. As a result we easily find

$$\begin{aligned} S = - \int \mathbf{e}_{fj} \cdot U [\nabla(u_{\mathbf{k}}^* u_{\mathbf{k}})] dV = - \frac{\hbar^2}{2m} \int \mathbf{e}_{fj} \cdot [\nabla u_{\mathbf{k}}^* \nabla^2 u_{\mathbf{k}} + \nabla u_{\mathbf{k}} \nabla^2 u_{\mathbf{k}}^*] dV - \\ - \frac{i\hbar^2}{m} \mathbf{k} \cdot \int (\nabla u_{\mathbf{k}}) (\mathbf{e}_{fj} \cdot \nabla u_{\mathbf{k}}^*) dV + \frac{i\hbar^2}{m} \mathbf{k}' \cdot \int u_{\mathbf{k}}^* (\mathbf{e}_{fj} \cdot \nabla u_{\mathbf{k}}) dV - \\ - \mathbf{e}_{fj} \cdot \int [E(\mathbf{k}) u_{\mathbf{k}} \nabla u_{\mathbf{k}}^* + E(\mathbf{k}') u_{\mathbf{k}}^* \nabla u_{\mathbf{k}}] dV + \\ + \frac{\hbar^2}{2m} k^2 \int u_{\mathbf{k}} \mathbf{e}_{fj} \cdot \nabla u_{\mathbf{k}}^* dV + \frac{\hbar^2 k'^2}{2m} \int u_{\mathbf{k}}^* \mathbf{e}_{fj} \cdot \nabla u_{\mathbf{k}} dV. \end{aligned} \quad (55.17)$$

The integration is carried out over a unit cell of the crystal, which is the basic region of periodicity of the functions $u_{\mathbf{k}}$ and $u_{\mathbf{k}}^*$. Hence on applying Green's formula the sum of the first two integrals is converted into surface integrals which reduce to zero.

The fifth and seventh terms can be integrated by parts, writing, taking into account the properties of periodicity,

$$\int u_{\mathbf{k}'}^* \nabla u_{\mathbf{k}} dV = - \int u_{\mathbf{k}} \nabla u_{\mathbf{k}'}^* dV.$$

Then we finally have

$$\begin{aligned} S = & -\frac{i\hbar^2}{m} \mathbf{k} \cdot \int \nabla u_{\mathbf{k}} (\mathbf{e}_{fj} \cdot \nabla u_{\mathbf{k}'}^*) dV + \frac{i\hbar^2}{m} \mathbf{k}' \cdot \int \nabla u_{\mathbf{k}'}^* (\mathbf{e}_{fj} \cdot \nabla u_{\mathbf{k}}) dV - \\ & - \mathbf{e}_{fj} \cdot \int \left[E(\mathbf{k}) - E(\mathbf{k}') + \frac{\hbar^2}{2m} (k^2 - k'^2) \right] u_{\mathbf{k}} \nabla u_{\mathbf{k}'}^* dV. \end{aligned} \quad (55.18)$$

The last integral contains the quantities $E(\mathbf{k}) - E(\mathbf{k}')$ and $(\hbar^2/2m)(k^2 - k'^2)$, which in order of magnitude are equal to the difference between the energies of the electron before and after scattering, i.e. are of the order of the phonon energy $\hbar\omega_f$. As we shall see below, the first integral is in order of magnitude equal to the kinetic energy of the electron. Hence it is substantially larger than the second integral which can be dropped.

Assuming in a rather rough approximation that $u_{\mathbf{k}} \approx u_{\mathbf{k}}^*$ and that these functions possess spherical symmetry, we can write the integral over an unit cell in the form

$$\mathbf{A} = \int_{\tau_0} \frac{\partial u_{\mathbf{k}}}{\partial r} \frac{\mathbf{r}}{r} \left(\mathbf{e}_{fj} \cdot \frac{\partial u_{\mathbf{k}}}{\partial r} \frac{\mathbf{r}}{r} \right) dV = \int_{\tau_0} \left(\frac{\partial u_{\mathbf{k}}}{\partial r} \right)^2 \frac{\mathbf{r}}{r^2} (\mathbf{e}_{fj} \cdot \mathbf{r}) dV. \quad (55.19)$$

In integrating, the vector \mathbf{r} runs over all possible values, hence it is clear that the vector \mathbf{A} can be directed only along the unique vector \mathbf{e}_{fj} . We then have

$$\mathbf{A} = \mathbf{e}_{fj} |\mathbf{A}|.$$

The modulus of the vector \mathbf{A} can be determined by multiplying it by the vector \mathbf{e}_{fj}

$$|\mathbf{A}| = \int_{\tau_0} \frac{1}{r^2} \left(\frac{\partial u_{\mathbf{k}}}{\partial r} \right)^2 (\mathbf{e}_{fj} \cdot \mathbf{r})^2 dV = \frac{1}{3} \int_{\tau_0} |\nabla u_{\mathbf{k}}|^2 dV. \quad (55.20)$$

After the transformations which have been carried out we can now write

the expression for S as

$$S = -\frac{i\hbar^2}{3m}(\mathbf{k} - \mathbf{k}') \cdot \mathbf{e}_{fj} \int_{\tau_0} |\nabla u_{\mathbf{k}}|^2 dV. \quad (55.21)$$

Setting $\mathbf{k}' = \mathbf{k} + \mathbf{f}$ and $\mathbf{k}' = \mathbf{k} - \mathbf{f}$, we obtain the expressions for S_+ and S_- given by formula (55.9).

§56. The total Hamiltonian of a solid

Making use of the results obtained above, we can write the total Hamiltonian of a solid in the form

$$\begin{aligned} H = & -\frac{\hbar^2}{2M} \sum \nabla_j^2 - \frac{\hbar^2}{2m} \sum \nabla_i^2 + \sum U(\mathbf{r}_i - \mathbf{R}_j) + \sum U(\mathbf{R}_i - \mathbf{R}_j) + \\ & + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} = \sum \hat{a}_{\mathbf{k},\sigma}^\dagger \hat{a}_{\mathbf{k},\sigma} \epsilon(\mathbf{k}) + \\ & + \sum \epsilon(f) (\hat{b}_f^\dagger \hat{b}_f + \frac{1}{2}) - i \sum_{\mathbf{k},\mathbf{f}} D_\omega (\hat{a}_{\mathbf{k}+\mathbf{f},\sigma}^\dagger \hat{a}_{\mathbf{k},\sigma} \hat{b}_f - \hat{a}_{\mathbf{k}-\mathbf{f},\sigma} \hat{a}_{\mathbf{k},\sigma}^\dagger \hat{b}_f^\dagger) + \\ & + \sum_{|\mathbf{r}_i - \mathbf{r}_j| > \lambda_{\min}} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{|\mathbf{r}_i - \mathbf{r}_j| < \lambda_{\min}} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}. \end{aligned} \quad (56.1)$$

Here $\epsilon(\mathbf{k})$ is the energy of the electron (quasiparticle) with wave vector \mathbf{k} . The operators $\hat{a}_{\mathbf{k},\sigma}^\dagger$ and $\hat{a}_{\mathbf{k},\sigma}$ represent the operators of creation and destruction of the electron (quasiparticle) with wave vector \mathbf{k} and spin index σ . The energy of the other quasiparticle, a phonon, is $\epsilon(\mathbf{f}) = \hbar\omega_f$.

The Coulomb interaction between electrons at large distances amounts to zero-point plasma oscillations, while at small distances it amounts to the screened interaction and the formation of quasiparticles. As to the interaction between a vibrating lattice and electrons (quasiparticles), it is formally described by the collision operator between two independent quasiparticles, a fermion and a boson. This interaction is the cause of transitions of the electron from one state into another.

The Hamiltonian of a solid in the second quantization representation was first obtained by Fröhlich. As an example of the application of the mathe-

mathematical technique of the second quantization method, we shall find an important formula for the probability of transition of an electron from state \mathbf{k} into state \mathbf{k}' with the emission or absorption of a phonon. In our problem the stationary states of the unperturbed system represent states with a definite number of phonons n_f and a definite number of electrons $n_{\mathbf{k}\sigma}$. The latter numbers, by virtue of the exclusion principle, are equal to one or zero.

If we confine ourselves to the first approximation of perturbation theory, then it turns out that only transitions with the absorption or emission of only one phonon are possible. Indeed, the perturbation operator involves the operators \hat{b}_f or \hat{b}_f^\dagger but not their products. The matrix elements of the operators \hat{b}_f and \hat{b}_f^\dagger differ from zero only for transitions in which the number of phonons changes by one (see (55.3)).

Let us for definiteness consider the case where the electron passes from a state with wave number \mathbf{k} and energy $E(\mathbf{k})$ to a state with wave number \mathbf{k}' and a larger energy. Then a phonon with wave number \mathbf{f} is absorbed, so that $\mathbf{k}' = \mathbf{k} + \mathbf{f}$. Since the total energy of the system (electrons+phonons) must remain constant, the energy of the electron after collision is

$$E(\mathbf{k}') = E(\mathbf{k}) + \hbar\omega_f.$$

In the second quantization representation, there corresponds to the process of transition a decrease by one in the number of electrons in state $\mathbf{k}\sigma$ and of the number of phonons in state \mathbf{f} and an increase by one of the number of electrons in state $\mathbf{k}'\sigma$.

To find the probability of such a transition, we have to calculate the matrix element

$$(H_{\text{int}})_{\mathbf{k}'\mathbf{k}} = \langle n_{\mathbf{k}'\sigma} = 1, n_{\mathbf{k}\sigma} = 0, n_f - 1 | H_{\text{int}} | n_{\mathbf{k}'\sigma} = 0, n_{\mathbf{k}\sigma} = 1, n_f \rangle. \quad (56.2)$$

Applying the rule for calculating the matrix elements of a product of operators and making use of the form of the matrix elements of $\hat{a}_{\mathbf{k}}^\dagger$, $\hat{a}_{\mathbf{k}}$, and \hat{b}_f^\dagger , \hat{b}_f , we easily find that of all the sum (56.1) only one matrix element remains. Namely,

$$(H_{\text{int}})_{\mathbf{k}'\mathbf{k}} = -iD_\omega \langle n_{\mathbf{k}'} = 1 | \hat{a}_{\mathbf{k}}^\dagger | n_{\mathbf{k}'} = 0 \rangle \langle n_{\mathbf{k}} = 0 | \hat{a}_{\mathbf{k}} | n_{\mathbf{k}} = 1 \rangle \langle n_f - 1 | \hat{b}_f | n_f \rangle. \quad (56.3)$$

Using formulae (99.15) of Part V, we see that the product of matrix elements of \hat{a}^\dagger , \hat{a} is equal to plus or minus one, depending on the occupation numbers. The matrix elements of the operator \hat{b}_f are defined by formula (47.17).

The probability of a transition with the absorption of a phonon is equal to

$$\begin{aligned} dW_- &= \frac{2\pi}{\hbar} |(H_{int})_{\mathbf{k}, \mathbf{k}'}|^2 \delta[E(\mathbf{k}) - E(\mathbf{k} + \mathbf{f}) + \hbar\omega_f] \frac{d\mathbf{f} dV}{(2\pi)^3} = \\ &= \frac{2\pi}{\hbar} D_\omega^2 n_f \frac{d\mathbf{f} dN}{(2\pi)^3} \delta[E(\mathbf{k}) - E(\mathbf{k} + \mathbf{f}) + \hbar\omega_f]. \end{aligned}$$

If we disregard dispersion and substitute the value of D_ω^2 from (55.12), we easily find

$$dW_- = \frac{4\pi g^2 n_f f}{9MNC} \delta\{E(\mathbf{k}) - E(\mathbf{k} + \mathbf{f}) + \hbar cf\} \frac{d\mathbf{f} dV}{(2\pi)^3}.$$

The total probability of the transition of an electron with the absorption of a phonon is obtained by integrating over all possible values of \mathbf{f} and V .

$$W_- = 4\pi \frac{g^2}{9MNC} \int f n_f \delta\{E(\mathbf{k}) - E(\mathbf{k} + \mathbf{f}) + \hbar cf\} \frac{d\mathbf{f} dV}{(2\pi)^3}. \quad (56.4)$$

The probability of a transition with the emission of a phonon can be found in exactly the same way:

$$W_+ = 4\pi \frac{g^2}{9MNC} \int f(n_f + 1) \delta\{-E(\mathbf{k}) + E(\mathbf{k} - \mathbf{f}) + \hbar cf\} \frac{d\mathbf{f} dV}{(2\pi)^3}. \quad (56.5)$$

Processes with the absorption or emission of one phonon can be presented



Fig. VI.10

in an obvious way in the form of a Feynman diagram (fig. VI.10). Here the wavy line represents the phonon, and the solid line stands for the electron. If one does not confine oneself to transitions of the first order, then other processes (shown in the Feynman diagrams of fig. VI.11 and fig. VI.12), are also possible.

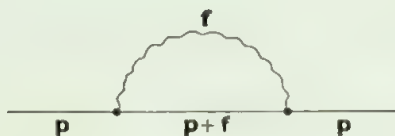


Fig. VI.11

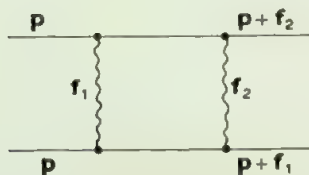


Fig. VI.12

An electron may emit a virtual phonon and absorb it on the spot. Because of this process the electron turns out to be surrounded by a cloud or jacket of virtual phonons with different frequencies. As the electron moves in a lattice the phonon cloud follows it, changing its mass. A quantitative calculation of this effect is beyond the scope of this book.

Another important second order effect is shown by the diagram in fig. VI.12. One of the electrons emits phonon f_1 and absorbs phonon f_2 . The other electron, on the contrary, absorbs the first phonon and emits the second phonon. This phonon exchange leads under certain conditions to a weak attraction between the electrons. In spite of the smallness of the effect, it plays a very important role in phenomena occurring in metals at low temperatures (see §65).

It should be noted that the above derivation of the interaction Hamiltonian involves the implicit assumption that, as a result of interaction with a phonon, the electron may always pass to a new state. This holds for electrons in the upper energy levels of the Fermi distribution.

In considering kinetic effects, for example electric conduction, we shall be

interested in the interaction with just these electrons. However, in studying some other properties of metals, such as the effect of the electrons on the energy spectrum of a system of phonons, it turns out to be essential to take into account the behaviour of the entire system of electrons, including the electrons of the filled band. These electrons cannot take part in transitions between states inside the band. Hence a description of effects involving the behaviour of the entire system of electrons cannot be obtained on the basis of the Hamiltonian (55.11).

The Kinetic Properties of Solids

§57. The kinetic equation for electrons in metals

The statement that physical kinetics has achieved its greatest successes in the field of solid-state theory and made it possible to understand and describe both qualitatively and quantitatively a very large number of diverse and detailed effects is hardly an exaggeration. Within the framework of this book we can consider only the basic results in this rapidly developing branch of physics.

We shall begin by considering kinetic phenomena in solids by finding the kinetic equation for charge carriers in metals. We have seen above that elementary excitations in a system of electrons in metals can be likened to a gas of free fermions. We shall call them conduction electrons and shall describe them by a one-particle distribution function $f(\mathbf{p}, \mathbf{r}, t)$. For brevity we shall call the quasi-momentum, \mathbf{p} , the momentum. This terminology cannot cause misunderstanding.

As conduction electrons move they undergo scattering on the vibrating atoms of the lattice (collisions with phonons) and on all kinds of inhomogeneities of the lattice which are called impurities. Sufficiently good results are obtained if this scattering is considered to be elastic (see §58). On the basis of what we said in §50, we shall disregard collisions between electrons.

If an external electromagnetic field is applied to the metal, then the dis-

tribution function satisfies the kinetic equation

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \dot{\mathbf{p}} \cdot \frac{\partial f}{\partial \mathbf{p}} = I. \quad (57.1)$$

The group velocity \mathbf{v} can be written in the form

$$\mathbf{v} = \frac{\partial \epsilon(\mathbf{p})}{\partial \mathbf{p}}. \quad (57.2)$$

Within the framework of the statistical description, the change in the quasi-momentum under the action of an external force can be taken into account in the quasi-classical approximation, i.e.

$$\dot{\mathbf{p}} = e \left(\mathcal{E} + \frac{1}{c} [\mathbf{v} \times \mathbf{H}] \right). \quad (57.3)$$

The collision integral for particles obeying Fermi statistics can be written in the approximation of perturbation theory.

The probability of a transition from state \mathbf{p} into state \mathbf{p}' can be written as

$$w_{\mathbf{p} \rightarrow \mathbf{p}'} = (1 - f(\mathbf{p}')) F(\mathbf{p}, \mathbf{p}', n_{\mathbf{p}}), \quad (57.4)$$

where the first factor gives the probability that state \mathbf{p}' is not occupied. The factor $F(\mathbf{p}, \mathbf{p}', n_{\mathbf{p}})$ represents the probability of a transition into the vacant state, due to all the processes of elastic interaction of the electrons with other particles; lattice phonons or impurities.

By virtue of the principle of microscopic reversibility, the probability for an inverse transition is of the form

$$w_{\mathbf{p}' \rightarrow \mathbf{p}} = (1 - f(\mathbf{p})) F(\mathbf{p}', \mathbf{p}, n_{\mathbf{p}}). \quad (57.5)$$

We shall come back to the formulation of the kinetic equation in §62, where the expression for the transition probability will be refined. For the present we shall leave out the dependence of $F(\mathbf{p}, \mathbf{p}', n_{\mathbf{p}})$ on the occupation numbers $n_{\mathbf{p}}$ and shall set

$$F(\mathbf{p}, \mathbf{p}', n_{\mathbf{p}}) = F(\mathbf{p}, \mathbf{p}').$$

Taking into account (57.4) and (57.5), the collision integral can be written in

the form

$$I = \int \{f(\mathbf{p}') (1 - f(\mathbf{p})) - f(\mathbf{p}) (1 - f(\mathbf{p}'))\} \delta(\epsilon - \epsilon') F(\mathbf{p}, \mathbf{p}') d\mathbf{p}', \quad (57.6)$$

as for collisions between electrons and ordinary atoms. Since the scattering is considered to be elastic, we can write

$$F(\mathbf{p}, \mathbf{p}') \delta(\epsilon - \epsilon') d\mathbf{p}' = F(\alpha) d\Omega, \quad (57.7)$$

where

$$\alpha = (\widehat{\mathbf{p}, \mathbf{p}'}).$$

Thus the kinetic equation assumes the form

$$\begin{aligned} \frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + e \left(\mathcal{E} + \frac{1}{c} [\mathbf{v} \times \mathbf{H}] \right) \cdot \frac{\partial f}{\partial \mathbf{v}} = \\ = \int \{f(\mathbf{p}') (1 - f(\mathbf{p})) - f(\mathbf{p}) (1 - f(\mathbf{p}'))\} F(\alpha) d\Omega. \end{aligned} \quad (57.8)$$

Relation (57.8) represents a linearized Boltzmann equation, very similar in its structure to eq. (27.4).

We note that for an uniform system in a stationary state, in the absence of external forces, eq. (57.8) turns into

$$f(\mathbf{p}) (1 - f(\mathbf{p}')) - f(\mathbf{p}') (1 - f(\mathbf{p})) = 0. \quad (57.9)$$

Taking into account isotropy, where the distribution function depends only on the absolute value of the momentum or, which is the same, on the energy, the solution of this functional equation is the Fermi distribution

$$f(\epsilon) = \frac{1}{e^{(\epsilon - \mu)/kT} + 1}. \quad (57.10)$$

The consideration of a large class of kinetic phenomena can be carried out without specifying the form of the function $F(\alpha)$. Hence in subsequent sections we shall assume $F(\alpha)$ to be a certain given function. We shall find the solution of the kinetic equation by the same methods as in the classical kinetic theory of gases. Then, making use of the expressions for the transition probabilities (56.9) and (56.10), we shall find the relaxation time or the free

path length. Finally, in order to convince ourselves of the correctness of these approximations, we shall, in §63, carry out a more consistent statement and solution of Boltzmann's equation for a metal.

In the case of weak fields and small departures from an equilibrium state, we shall, as in §27, try to find the solution of (57.8) in the form

$$f = f_0(\epsilon) + f_1(\mathbf{p}, \Omega, t), \quad (57.11)$$

where $f_0 \gg f_1$.

Then for f_1 we obtain an equation which is the same as eq. (27.10):

$$\begin{aligned} \frac{\partial f_1}{\partial t} + \mathbf{v} \cdot \left\{ (\epsilon - \mu) \nabla \left(\frac{1}{kT} \right) - \frac{\nabla \mu}{kT} + e \mathcal{E} \right\} \frac{\partial f_0}{\partial \epsilon} + \\ + \frac{e}{c} [\mathbf{v} \times \mathbf{H}] \cdot \frac{\partial f_1}{\partial \mathbf{p}} + \mathbf{v} \cdot \frac{\partial f_1}{\partial \mathbf{r}} = \int (f'_1 - f_1) F(\alpha) d\Omega', \end{aligned} \quad (57.12)$$

where \mathcal{E} is the electric field.

§58. The electrical conductivity of metals

If a constant electric field \mathcal{E} is applied to a metal, then a stationary current will exist in it. It is necessary to have a clear idea as to why the motion of conduction electrons turns out to be stationary. If the collision integral in eq. (57.8) could be dropped, then the equation for homogeneous species would assume the form

$$\frac{\partial f}{\partial t} + e \mathcal{E} \cdot \frac{\partial f}{\partial \mathbf{p}} = 0$$

and the distribution function would depend on time. In fact this would mean that the electrons were performing an accelerated motion under the action of the force $e\mathcal{E}$; their velocity, and hence also the current, would increase continuously with time.

The origin of the momentum loss compensating for the accelerating action of the field is in collisions between the electrons and phonons or impurities.

Boltzmann's equation for the time-independent distribution function can be written only by taking into account the collision integral.

Without reproducing the calculations of §27, we can write for the current

density the expression

$$\mathbf{j} = e \int \mathbf{v} f_1 d\mathbf{p} = -\frac{e^2}{m} \int \lambda_{tr} \frac{\partial f_0}{\partial \epsilon} \frac{\mathbf{p}}{v} \cdot (\mathbf{v} \cdot \mathcal{E}) d\mathbf{p}. \quad (58.1)$$

Thus the electrical conductivity turns out to be equal to

$$\sigma_{ik} = -\frac{e^2}{m} \int \lambda_{tr} \frac{v_i}{v} p_k \frac{\partial f_0}{\partial \epsilon} d\mathbf{p}. \quad (58.2)$$

In crystals with cubic symmetry the field \mathcal{E} and current \mathbf{j} are parallel to each other. Choosing the direction of the vector \mathcal{E} to be the x -axis, one can rewrite (58.1) in the form

$$j = -e^2 \mathcal{E} \int \lambda_{tr} \left(\frac{\partial f_0}{\partial \epsilon} \right) v \cos^2 \theta d\mathbf{p}, \quad (58.3)$$

where θ is the angle between the vectors \mathbf{v} and \mathcal{E} .

In formula (58.3) it is convenient to change to integration over energies and angles. This gives

$$\sigma = \frac{16\pi e^2 m}{3(2\pi\hbar)^3} \int_0^\infty \lambda_{tr} \epsilon \frac{\partial f_0}{\partial \epsilon} d\epsilon. \quad (58.4)$$

We now make use of the property of the function $\partial f_0 / \partial \epsilon$ for the Fermi distribution in a metal considered in §80 of Part III.

We have, analogously to (80.5) of Part III,

$$L_1 = \int \lambda_{tr} \epsilon \frac{\partial f_0}{\partial \epsilon} d\epsilon \approx \int_{-\mu/kT}^\infty \lambda_{tr}(\mu + kTx) \frac{\partial f_0}{\partial x} dx \approx \int_{-\infty}^\infty \lambda_{tr}(\mu + kTx) \frac{\partial f_0}{\partial x} dx,$$

where $x = (\epsilon - \mu)/kT$.

If the free path length λ_{tr} depends on the energy, then, since according to the data of §80 of Part III the derivative $\partial f_0 / \partial x$ behaves as a δ -function, we have

$$L_1 = \int_{-\infty}^\infty \lambda_{tr}(\epsilon) \frac{\partial f_0}{\partial x} (\mu + kTx) dx \approx -\lambda_{tr}(\mu) \mu. \quad (58.5)$$

Hence for the electrical conductivity we obtain the expression

$$\sigma = \frac{j_z}{E} = \frac{16\pi e^2 m}{3(2\pi\hbar)^3} \lambda_{\text{tr}}(\mu) \mu, \quad (58.6)$$

or, on the basis of (80.10) of Part III,

$$\sigma = \frac{16\pi e^2 m \lambda_{\text{tr}}(\mu) \epsilon_{\text{max}}}{3(2\pi\hbar)^3} = \frac{16\pi e^2 m \lambda_{\text{tr}}(\mu)}{3(2\pi\hbar)^3} \frac{\hbar^2}{2m} \left(\frac{3}{8\pi}\right)^{\frac{2}{3}} \left(\frac{N}{V}\right)^{\frac{2}{3}}. \quad (58.7)$$

This final formula can also be written in another form by introducing the velocity $v(\mu)$ of an electron having energy $\epsilon = \mu$. By means of (79.3) of Part III we find

$$\sigma = \frac{e^2 \lambda_{\text{tr}}(\mu)}{m v(\mu)} \frac{N}{V} = \frac{e^2 \lambda_{\text{tr}}(\mu)}{m v(\mu)} n. \quad (58.8)$$

We see that the electrical conductivity depends on the number of electrons per unit volume, n , as well as on the velocity of an electron at the Fermi surface, $v(\mu)$, and on the mean free path of an electron having an energy on the Fermi surface.

The first important feature of the expression obtained for the electrical conductivity of a metal is the fact that it turns out not to be proportional to the number of electrons per unit volume; in formula (58.8) the velocity, $v(\mu)$, is expressed in terms of n . The cause of this is clear: only electrons whose states lie in the upper levels of the Fermi distribution take part in electrical conduction. Only these electrons are conduction electrons. Electrons in states lying in the filled band cannot perform a systematic motion and carry current.

Formula (58.8) for the electrical conductivity contains two unknown quantities, the number of electrons per unit volume n , and the mean free path of electrons having an energy lying on the Fermi surface, $\lambda(\mu)$. As we have already said, the velocity of electrons $v(\mu)$ is expressed in terms of n by the formulae of §80 of Part III.

In the next section it will be shown that the quantity n can be defined independently of resistance. Then (58.8) allows the mean free path to be expressed in terms of the electrical conductivity σ . The value of the latter can be measured directly. It turns out that the value of λ_{tr} substantially exceeds the interatomic distance (the lattice constant). Furthermore, λ_{tr} depends strongly on the temperature. For example, for copper, for which it can be assumed that there is one free electron to each atom of the lattice, λ_{tr} varies

from 7×10^{-7} cm at $T \approx 1300$ K up to 4×10^{-5} cm at $T \approx 100$ K, which is larger by a factor of 300–2000 than the distance between atoms in the lattice.

It turns out that the temperature dependence of λ_{tr} , and with it of the electrical conductivity σ , differs for the cases of temperatures which are high and low relative to the Debye temperature Θ .

At high temperatures $T \gg \Theta$, $\sigma \sim T$. At low temperatures σ shows a rapid increase, proportional to $1/T^5$, with decreasing temperature. At $T = 0$ in a perfectly pure metal $\sigma \rightarrow \infty$, and the ohmic resistance tends to zero.

In samples containing impurities, as well as in samples with mechanical defects, dislocations, residual mechanical stresses, etc., at very low temperatures (of the order of a few degrees Kelvin) there is a residual resistance which does not depend on the temperature and is proportional to the concentration of impurities.

Thus the electrical conductivity of a metal at $T \neq 0$ has a finite value, and the resistance differs from zero. Only at $T = 0$ and in the case of a sample containing no impurities and having no defects does a metallic conductor show no resistance to the flow of current. Such a conductor is said to be ideal. Since it is impossible to obtain a temperature equal to zero, and since it is also impossible to produce a completely pure sample, the ideal conductor represents a limit to which one can aspire by lowering the temperature and purifying the metal.

§59. The Hall effect

Important information about the properties of an electron gas can be obtained by studying the behaviour of metals placed in an external magnetic field.

If a metallic sample in which a current flows (the direction of which is taken to be the x -axis) is placed in a magnetic field directed along the z -axis, then an electric field E_y will arise in the sample.

The origin of this effect, called the Hall effect, is quite clear.

Under the action of the Lorentz force the electrons which form the current are deflected in the magnetic field into the negative direction of the y -axis. They accumulate at the lower face of the metal until the electric field they produce compensates for the action of the deflecting force. The electrons subsequently are in a steady state.

From the above it is clear that the direction of the field E_y is defined by the sign of the charge carriers. We shall also see that its magnitude is directly

related to the number of charge carriers per unit volume. These two facts make the Hall effect one of the most important methods of studying the properties of metals and, as will be seen in the following, semiconductors. Let us turn to the calculation of the field E_y .

We write Boltzmann's kinetic equation, taking into account that in our case the force F has the components

$$F_x = eE_x + \frac{e}{c} H v_y ,$$

$$F_y = eE_y - \frac{e}{c} H v_x ,$$

$$F_z = 0 .$$

Then the kinetic equation assumes the form

$$e \left(E_x + \frac{v_y}{c} H \right) \frac{\partial f}{\partial p_x} + e \left(E_y - \frac{v_x}{c} H \right) \frac{\partial f}{\partial p_y} = I . \quad (59.1)$$

As in §27, we shall follow the Lorentz method and seek the solution of Boltzmann's equation (59.1) in the form

$$f(\mathbf{p}) = f_0(p_0) + p_x f_1(p) + p_y f_2(p) , \quad (59.2)$$

since there are electric field components in the directions of the x - and y -axes. We then obtain

$$I = - \frac{vp}{\lambda_{tr}} (f_1 \cos \theta + f_2 \sin \theta) , \quad (59.3)$$

where θ is the angle between the direction of the momentum and the x -axis.

Substituting the expansion (59.2) into (59.1), we find to within the second order of small quantities:

$$e \left(E_x + \frac{v_y H}{c} \right) \left(\frac{\partial f_0}{\partial p_x} + p_x \frac{\partial f_1}{\partial p_x} + f_1 \right) =$$

$$\begin{aligned}
&= e \left(E_x + \frac{v_y H}{c} \right) \left(\frac{\partial f_0}{\partial \epsilon} v_x + f_1 + p_x \frac{\partial f_1}{\partial p_x} \right) \approx \\
&\approx e \left(E_x v_x \frac{\partial f_0}{\partial \epsilon} + \frac{v_x v_y H}{c} \frac{\partial f_0}{\partial \epsilon} + F_x f_1 + E_x p_x \frac{\partial f_1}{\partial p_x} + \right. \\
&\quad \left. + v_y \frac{H}{c} f_1 + \frac{v_y p_x H}{c} \frac{\partial f_1}{\partial p_x} \right) \approx c \left(E_x v_x \frac{\partial f_0}{\partial \epsilon} + \frac{v_y H}{c} f_1 \right).
\end{aligned}$$

The terms of the second order of small quantities $E_x p_x \partial f_1 / \partial p_x$, $(v_y p_x / c) \times \partial f_1 / \partial p_x$ and higher orders are omitted. The term $(v_y H / c) f_1$ cannot be dropped, as will be seen from subsequent calculations. Analogously we find

$$\begin{aligned}
&e \left(E_y - \frac{v_x H}{c} \right) \left(\frac{\partial f_0}{\partial p_y} + f_2 + p_y \frac{\partial f_2}{\partial p_y} \right) \approx \\
&\approx e \left(E_y v_y \frac{\partial f_0}{\partial \epsilon} - \frac{v_x v_y}{c} H \frac{\partial f_0}{\partial \epsilon} - \frac{v_x H}{c} f_2 + \dots \right) \\
&\approx e E_y v_y \frac{\partial f_0}{\partial \epsilon} - \frac{e v_x H}{c} f_2.
\end{aligned}$$

Hence the left-hand side of (59.1) assumes the form

$$ve \left(E_x \frac{\partial f_0}{\partial \epsilon} - \frac{H}{c} f_2 \right) \cos \theta + ve \left(E_y \frac{\partial f_0}{\partial \epsilon} + \frac{H f_1}{c} \right) \sin \theta.$$

Equating this expression with the collision integral (59.3) and combining all terms containing $\cos \theta$ and $\sin \theta$, we find

$$\begin{aligned}
&\left[ve \left(E_x \frac{\partial f_0}{\partial \epsilon} - \frac{1}{c} H f_2 \right) + \frac{vp}{\lambda_{tr}} f_1 \right] \cos \theta + \\
&+ \left[ve \left(E_y \frac{\partial f_0}{\partial \epsilon} + \frac{H}{c} f_1 \right) + \frac{vp}{\lambda_{tr}} f_2 \right] \sin \theta = 0.
\end{aligned}$$

In view of the arbitrariness of the angle θ , we obtain two equations defin-

ing the functions f_1 and f_2 sought:

$$\frac{e}{m} E_x \frac{\partial f_0}{\partial \epsilon} - \frac{eH}{mc} f_2 = -\frac{v}{\lambda_{tr}} f_1 ,$$

$$\frac{e}{m} E_y \frac{\partial f_0}{\partial \epsilon} + \frac{eH}{mc} f_1 = -\frac{v}{\lambda_{tr}} f_2 .$$

Solving this system gives

$$f_1 = -\frac{\lambda_{tr} e}{vm} \frac{\partial f_0}{\partial \epsilon} \left(E_x + \frac{\lambda_{tr} \omega_H}{v} E_y \right) / \left[1 + \left(\frac{\lambda_{tr} \omega_H}{v} \right)^2 \right] ,$$

$$f_2 = -\frac{\lambda_{tr} e}{vm} \frac{\partial f_0}{\partial \epsilon} \left(E_y - \frac{\lambda_{tr} \omega_H}{v} E_x \right) / \left[1 + \left(\frac{\lambda_{tr} \omega_H}{v} \right)^2 \right] ,$$

where

$$\omega_H = eH/mc .$$

Knowing the corrections to the distribution function, one can find the current in the direction of the x -axis and the field E_y .

Namely, we have

$$j_x = e \int v_x f d\mathbf{p} = -C(E_x L_1 - E_y L_2) , \quad (59.4)$$

$$j_y = 0 = e \int v_y f d\mathbf{p} = C(E_y L_1 + E_x L_2) , \quad (59.5)$$

where

$$L_1 = \int \frac{\lambda_{tr} e}{1 + (\lambda_{tr} \omega_H/v)^2} \frac{\partial f_0}{\partial \epsilon} d\epsilon , \quad (59.6)$$

$$L_2 = \int \frac{\lambda_{tr} e (\lambda_{tr} \omega_H/v)}{1 + (\lambda_{tr} \omega_H/v)^2} \frac{\partial f_0}{\partial \epsilon} d\epsilon , \quad (59.7)$$

$$C = \frac{16\pi m e^2}{3(2\pi\hbar)^3} . \quad (59.8)$$

For brevity we drop the subscript tr of λ . It is obvious that for $H \rightarrow 0$

$$L_1 \rightarrow \int \lambda e \frac{\partial f_0}{\partial \epsilon} d\epsilon \quad \text{and} \quad L_2 \rightarrow 0.$$

In this case j_x goes over into (58.3).

To calculate L_1 and L_2 use can be made of the properties of the function $\partial f_0 / \partial \epsilon$. In the first approximation

$$L_1 = -\lambda(\mu) \mu \left[1 + \left(\frac{\mu(\mu) \omega_H}{v(\mu)} \right)^2 \right]^{-1}, \quad (59.9)$$

$$L_2 = -\frac{\lambda^2(\mu) \mu \omega_H}{v(\mu)} \left[1 + \left(\frac{\lambda(\mu) \omega_H}{v(\mu)} \right)^2 \right]^{-1} = \frac{\lambda(\mu) \omega_H}{v(\mu)} L_1, \quad (59.10)$$

where $\lambda(\mu)$ and $v(\mu)$ are taken at the Fermi surface.

From formulae (59.4) and (59.5) it is possible to find the conductivity of a metal placed in a magnetic field, $\sigma(H) = j_x / E_x$, and the transverse field E_y .

Let us first find this last quantity:

$$\begin{aligned} E_y &= \frac{L_2}{C(L_1^2 + L_2^2)} j_x = \frac{L_2}{L_1} \frac{j_x}{L_1 C} \frac{1}{1 + (L_2/L_1)^2} = \\ &= \frac{\lambda(\mu) \omega_H}{C v(\mu)} \frac{[1 + (\lambda(\mu) \omega_H / v(\mu))^2]}{\lambda(\mu) \mu} j_x [1 + (\lambda(\mu) \omega_H / v(\mu))^2]^{-1} \\ &= \frac{\omega_H}{C v(\mu) \mu} j_x = \frac{e H j_x}{m c \mu v(\mu) C}. \end{aligned} \quad (59.11)$$

Substituting here the value $\mu \approx \epsilon_{\max}$ and $v(\mu)$ and C from (59.8), we obtain after some simple transformations

$$E_y = R H j_x, \quad (59.12)$$

where R denotes the quantity

$$R = \frac{1}{e n c}. \quad (59.13)$$

Formula (59.12) shows that the transverse field is proportional to the current j_x and to the strength of the magnetic field H .

The factor of proportionality R , called the Hall constant, depends, as is seen from its definition, on only two quantities, the charge of current carrier e and the number of particles per unit volume. The sign of R and, consequently, also the direction of the transverse field E_y , are necessarily the same as the sign of the current carrier.

The conductivity in a magnetic field is

$$\sigma(H) = C \frac{L_1^2 + L_2^2}{L_1} = \frac{L_2}{L_1} \frac{1}{RH} \approx \frac{\lambda(\mu)e}{mcv(\mu)} \frac{1}{R}. \quad (59.14)$$

Since R does not, in the first approximation depend on the strength of the magnetic field, (as regards calculation of the integrals L_1 and L_2) the right-hand side does not depend on H . This means that in this approximation $\sigma(H) = \sigma$, i.e. no change of resistance arises in a magnetic field. In higher approximations the conductivity $\sigma(H)$ turns out to depend on the field strength.

A set of measurements of the conductivity σ and of the Hall constant makes it possible to find the two unknown quantities: n and λ . Conversely, by defining the number of free electrons per atom one can calculate R .

The sign of the Hall constant is negative when the transport is carried out by electrons. This applies to monovalent metals.

In the case of divalent metals of the transition groups, where band overlap takes place, holes as well as electrons take part in the conductivity. Hence the sign often turns out to be positive. Anisotropic behaviour of the Hall constant, particularly pronounced in the case of metals such as Bi, is also observed.

We shall come back to the question of the sign of the Hall effect in §68.

In the next approximation $\sigma(H)$ turns out to be inversely proportional to the square of the magnetic field strength in a relatively weak field and tends to a constant value in a very strong field. However, in strong fields the behaviour of $\sigma(H)$ and its numerical value are in poor agreement with experimental data. This is associated with the crudeness of the model used above. Taking into account the effects of anisotropy, which always occur in real crystals, the agreement of theory with experiment can be considerably improved.

§60. The optical properties of a system of conduction electrons

We shall base the treatment of the optical properties of a metal on the as-

sumption that the interactions of the electromagnetic field of a light wave with the conduction electrons and with the electrons of the atom cores take place independently. The interaction of atoms with an electromagnetic field was discussed in Part V.

Therefore we shall confine ourselves to a discussion of the behaviour of a system of conduction electrons in the field of a light wave. We note that if the frequency, ω , of the field is not the same as one of the natural atomic frequencies, then the basic optical characteristics of the metal are determined only by the behaviour of the conduction electrons. In this case it is simplest to find the complex conductivity as a function of the frequency of the field acting on the metal.

As in calculating the electrical conductivity, we shall make use of Boltzmann's kinetic equation (58.12), which in our case takes the form

$$e\mathcal{E} \cdot \frac{\partial f_0}{\partial \mathbf{p}} + \mathbf{v} \cdot \frac{\partial f_1}{\partial \mathbf{r}} + \frac{\partial f_1}{\partial t} = \int (\mathcal{U}'_1 - f) F(\alpha) d\Omega' \simeq \frac{f'_1 - f}{\tau}. \quad (60.1)$$

We have retained in it the term $\partial f_1 / \partial \mathbf{r}$, but dropped the term with the magnetic field (since it is small with respect to v/c).

Suppose the external field varies according to

$$\mathcal{E} = \mathcal{E}_0 e^{i(\mathbf{\kappa} \cdot \mathbf{r} - \omega t)}.$$

Then it is natural to try to find the solution of eq. (60.1) in the form

$$f_1 = \alpha(\mathbf{v} : \mathcal{E}_0) e^{i(\mathbf{\kappa} \cdot \mathbf{r} - \omega t)}. \quad (60.2)$$

Substituting into (60.1) gives

$$e\mathcal{E}_0 \cdot \mathbf{v} \frac{\partial f_0}{\partial \mathbf{p}} = \alpha \mathbf{v} \cdot \mathcal{E}_0 \left[i(\mathbf{\kappa} \cdot \mathbf{v}) - i\omega + \frac{1}{\tau} \right].$$

Hence

$$f_1 = \frac{e\mathbf{v} \cdot \mathcal{E}_0 (\partial f_0 / \partial \mathbf{p}) e^{i(\mathbf{\kappa} \cdot \mathbf{r} - \omega t)}}{\tau^{-1} - i\omega + i(\mathbf{\kappa} \cdot \mathbf{v})}. \quad (60.3)$$

Since $\kappa \sim \omega/c$ and $v \ll c$, the term arising from the derivative $\partial f_1 / \partial t$ is small

and one can write

$$f_1 \simeq \frac{e\mathbf{v} \cdot \mathcal{E}_0 (\partial f_0 / \partial \epsilon) v^{-1} \lambda_{tr} e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}}{1 - i\omega\tau}. \quad (60.4)$$

We see that the correction to the distribution function turns out to be complex. Substituting (60.4) into (58.1) and reproducing the calculations of §58, it is easy to find for the conductivity the expression

$$\sigma(\omega) = \sigma(0) \frac{1 + i\omega\tau}{1 + \omega^2\tau^2}, \quad (60.5)$$

where $\sigma(0)$ is the conductivity in a constant field. Since τ is known from the electrical conductivity in a constant field, formula (60.5) gives a complete quantitative description of the optical properties of a system of conduction electrons. The formula for conductivity in a high-frequency field is valid in the region of the skin depth, which is given by formula (30.4) of Part IV. It is necessary that the skin depth δ be large in comparison with the mean free path λ_{tr} .

§61. The photoelectric effect

When a metal is irradiated by light of a sufficiently high frequency photoelectrons are emitted from its surface. This phenomenon, called the photoelectric effect, is well investigated and has found well-known practical applications. However, a complete and consistent theory of the photoelectric effect has only recently* been developed. This theory serves as an illustration of a certain general approach to the study of the threshold phenomena in quantum mechanics. These threshold phenomena are transitions at energies close to the energy threshold of a given process (see §93 of Part V).

Let us consider a metal contiguous to an optically transparent medium. Such a medium could be a solid dielectric, a solution or vacuum. When the surface of the metal is irradiated by light of frequency ω , a one-phonon transition with the emission of an electron from the metal can occur only for $\omega > \omega_0$, where the threshold frequency ω_0 is connected with the work function W by the obvious relation $\omega_0 = W/\hbar$. We shall confine ourselves to fre-

* The treatment in this section is based on a study of A.M. Brodsky, Yu. Ya. Gurevich and B.G. Levich, *Sov. Phys. Solid State* 40 (1970) 139.

quencies ω close to ω_0 and shall seek the probability of emission of photoelectrons as a function of the frequency ω of the light.

The photoelectric effect in metals may in principle have two different mechanisms:

(1) The so-called surface photoelectric effect in which photons collide with electrons in the surface layer of the metal. The surface layer is taken to be the region in which the potential energy varies from the value W inside the metal down to zero at its surface. An electron in the surface layer is in a field of force varying from point to point. This ensures fulfillment of the laws of conservation of energy and momentum in a collision between an electron and a photon accompanied by the emission of an electron from the metal.

(2) The volume photoelectric effect, taking place in the region of optical transparency of metals. This region usually lies in the ultraviolet part of the spectrum.

In the volume photoelectric effect the interaction of photons with electrons takes place inside the metal (in the region of constant potential energy). The role of the third body ensuring the fulfillment of conservation laws is in this case played by phonons or impurity atoms.

For photon energies lower than 8–10 eV the metals are not optically transparent and only the surface photoelectric effect may take place. The alkali metals, which are optically transparent in the visible region are an exception to this.

We shall confine ourselves to a treatment of the surface photoelectric effect. In this case the energy, E , of the photoelectrons emitted, is small in comparison with $\hbar\omega$, so that the photoelectrons can be considered to be slow particles. We note that the energy of photoelectrons usually amounts to about 1 eV, so that it is still large compared to kT .

We shall assume that the metal occupies the half-space $-\infty \leq z \leq 0$ and that it has an uniform surface. The electron moves in a potential well with potential energy $U(z)$, which for sufficiently large values of z goes over into a strictly periodic field inside the crystal. The mean value of the potential energy inside the metal is equal to

$$U(z) = -W = \text{const.} \quad (61.1)$$

In the region $0 \leq z \leq \delta$, i.e. in the surface layer of the medium adjacent to the metal, the electron is in a complex, unknown, potential field.

We shall assume the width of the layer δ to be small, so that the following inequality is fulfilled:

$$p\delta/\hbar \ll 1, \quad (61.2)$$

where p is the momentum of an electron emitted in the direction of the z -axis. In other words, we shall consider the electron wavelength $\lambda \ll \delta$. Finally, for $z > \delta$ the potential energy in the medium can be written as

$$U(z) = -V_0 - \frac{e^2}{4\epsilon_{\text{eff}}z}. \quad (61.3)$$

The first term has the meaning of the potential energy of an electron in a medium at a large distance from the surface (we assume its energy in vacuum to be equal to zero), and the second term represents the image force in a medium with effective dielectric constant ϵ_{eff} . Since the electron moves in the medium with an energy considerably larger than the thermal energy, ϵ_{eff} is not the same as the static dielectric constant but approaches rather the optical dielectric constant.

Let us write the Schrödinger equation for an electron inside and outside a metal. Since the metal is assumed to be uniform in the xy -plane, the unperturbed wave function can be written in the form

$$\psi = \psi_0(E, p, z) e^{i\mathbf{p}_{\parallel} \cdot \mathbf{p}}$$

where \mathbf{p}_{\parallel} and \mathbf{p} are the momentum and radius vector in the xy -plane, E is the energy of the electron, and p is the momentum along the z -axis.

The unperturbed Schrödinger equation has the form

$$\left\{ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + \frac{p_{\parallel}^2}{2m} - E + U(z) \right\} \psi_0 = 0. \quad (61.4)$$

When a radiation field which can be characterized by the operator $U_A(z)e^{i\omega t}$ is applied, one can write in the first approximation of perturbation theory

$$\psi = \psi_0 + \psi'$$

where ψ' satisfies the equation

$$\left\{ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + U(z) - (E + \hbar\omega) + \frac{p_{\parallel}^2}{2m} \right\} \psi' = -U_A(z) \psi_0. \quad (61.5)$$

Here $(E + \hbar\omega)$ is the energy of an electron which has absorbed a photon.

Making use of the energy conservation law

$$E + \hbar\omega = \frac{p^2 + p_{\parallel}^2}{2m} - V_0,$$

we rewrite (61.5) in the form

$$\left\{ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} - \frac{p_{\parallel}^2}{2m} + U(z) + V_0 \right\} \psi' = -U_A(z) \psi_0. \quad (61.6)$$

Let us consider the solution of eq. (61.6) in the regions $z > \delta$ and $z < 0$. In the first region, taking into account the rapid fall of ψ_0 outside the metal, the term $U_A \psi_0$ on the right-hand side of (61.6) can be dropped. The energy conservation law

$$\begin{aligned} E &= \frac{p^2}{2m} + \frac{p_{\parallel}^2}{2m} + U(z) - \hbar\omega = \\ &= \frac{p^2}{2m} + \frac{p_{\parallel}^2}{2m} - V_0 - \frac{e^2}{\epsilon_{\text{eff}} z} - \hbar\omega = \text{const.} \end{aligned}$$

allows one to write (61.6) in the form

$$\left\{ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} - \frac{p_{\parallel}^2}{2m} + \frac{e^2}{4\epsilon_{\text{eff}} z} \right\} \psi'(z) = 0. \quad (61.7)$$

Eq. (61.7) is the same as the equation describing the motion of an electron in a one-dimensional Coulomb field. The asymptotic expression we need for ψ' in the Coulomb field when $z \rightarrow \infty$, corresponding to a particle going off to infinity, has the form

$$\psi' = \chi_{z \rightarrow \infty}^+ \sim \exp [i(pz - \eta \ln(2pz/\hbar) + \delta_0)] , \quad (61.8)$$

where

$$\eta = 2me^2/\epsilon_{\text{eff}} p, \quad (61.9)$$

$$\delta_0 = \arg \Gamma(1 - i\eta). \quad (61.10)$$

When $z \rightarrow 0$, one can use for ψ' , formula (38.6) of Part V, setting

$$\psi' \sim \chi_{z \rightarrow 0}^+ \sim \left[\frac{1 - e^{-2\pi\eta}}{2\pi\eta} \right]^{\frac{1}{2}} \frac{pz}{\hbar}. \quad (61.11)$$

Hence $\psi'(z)$ can be written in the form

$$\psi'(z) = C(p, \omega) \chi(pz/\hbar, \eta), \quad (61.12)$$

where χ is the Coulomb function with the asymptotic behaviour mentioned, and $C(p, \omega)$ is a constant independent of z . To define it, (61.12) must be matched with the solution inside the metal. As was done in §93 of Part V, use must be made of the condition (61.2). Instead of carrying out the matching for $z = \delta$, we shall match (61.12) with the solution inside the metal for $z = 0$. The wave function of a particle with a wavelength $\lambda \gg \delta$ has no time to change substantially over the length δ . Examining the region $z < 0$, we note that inside the metal the potential energy $U(z) + W$ is very large in comparison with the quantity $p^2/2m$. Hence for $z < 0$ the term $p^2/2m$ in the Schrödinger equation can be dropped. As a result the wave function inside the metal turns out to be independent of the quantity p .

At the point $z = 0$ one has to write the condition for matching the function (61.12), which depends in general on p , with a function independent of p . It is clear that the condition can be fulfilled for all values of p , if for $z \rightarrow 0$ the ψ' defined by (61.12) also does not depend on p .

For this one has to set

$$C(p, \omega) = \left[\frac{2\pi\eta}{1 - e^{-2\pi\eta}} \right]^{\frac{1}{2}} f(\omega), \quad (61.13)$$

where $f(\omega)$ is a function independent of p . Thus for $z \rightarrow \infty$ we obtain

$$\psi' \sim \left[\frac{2\pi\eta}{1 - e^{-2\pi\eta}} \right]^{\frac{1}{2}} e^{ipz/\hbar}. \quad (61.14)$$

Knowing the wave function, one can find the photoelectric current. Namely, the probability current density is

$$j = \frac{\hbar}{2mi} (\psi^* \nabla \psi - \psi \nabla \psi^*) = \left(\frac{2E_0}{m} \right)^{\frac{1}{2}} [1 - \exp(-(E_0/E)^{\frac{1}{2}})]^{-1} f(\omega), \quad (61.15)$$

where $E_0 = 33.5/\epsilon_{\text{eff}}^2$ eV.

The total photoelectric current from unit surface of the metal is given by the formula

$$I = e \int j f(E) \rho(E, \mathbf{p}_{\parallel}) \theta(p) dE d\mathbf{p}_{\parallel} =$$

$$\begin{aligned}
&= e \left(\frac{2E}{m} \right)^{\frac{1}{2}} \int_{W'}^{\infty} [1 + e^{(E-\mu)/kT}]^{-1} [1 - e^{-(E_0/E)^{\frac{1}{2}}}]^{-1} dE \times \\
&\times \int_0^{[2m(E+V-\hbar\omega)]^{\frac{1}{2}}} \rho(E, \mathbf{p}_{\parallel}) 2\pi |\mathbf{p}_{\parallel}| d|\mathbf{p}_{\parallel}| \quad (61.16)
\end{aligned}$$

where $f(E)$ is the Fermi distribution, and $\rho(E, \mathbf{p}_{\parallel})$ is the number of states with given E and \mathbf{p} . The chemical potential μ is connected with the work function and the potential V_0 by the obvious relation $\mu = -(W - V_0)$.

Assuming $\rho(E, \mathbf{p}_{\parallel})$ to be a slowly varying function and taking it out of the integral sign, we have

$$I = \rho(E, \mathbf{p}_{\parallel}) e \int_{W'}^{\infty} j(E) f(E) dE \int 2\pi |\mathbf{p}_{\parallel}| d|\mathbf{p}_{\parallel}|. \quad (61.17)$$

Calculations for the two cases of emission of electrons into a vacuum and into a medium lead to the following results.

In the first case the energy of the photoelectrons is $E < E_0$, and

$$I \sim (\omega - \omega_0)^2. \quad (61.18)$$

This law for the dependence of photoelectric current on frequency (Fowler's law) is in good agreement with experiment. On the other hand when electrons are emitted into a dielectric or into a solution of an electrolyte for which $\epsilon_{\text{eff}} \sim 10$, $E > E_0$ and $j \sim E^{\frac{1}{2}}$. In this case

$$I \sim (\omega - \omega_0)^{\frac{5}{2}}. \quad (61.19)$$

This frequency dependence has been well confirmed by experimental data for the photocurrent at a metal-solution interface.

§62. The mean free path of electrons in metals

As in §27, in the approximation (57.11) one can introduce the concept of the mean free path λ_{tr}

$$\lambda_{tr} = \frac{1}{N\sigma_{tr}} = \frac{v_{el}}{W},$$

where v_{el} is the mean velocity of the electron, and W is the probability of its collision (per unit time) with the scatterer. The quantities v_{el} and W have a direct meaning. Of course, the mean free path cannot be understood literally as the distance between successive collisions.

As has already been mentioned above, in a metal collisions of electrons with phonons and with impurities in the lattice may occur. We shall denote the corresponding mean free paths by λ_{ph} and λ_{imp} . It is obvious that the electrical conductivity and other kinetic quantities are determined by the value of the smaller of these two mean free paths (i.e. by the more probable collision). As will be seen from subsequent calculations, this mean free path is usually λ_{ph} . Hence we shall restrict our considerations with a calculation of λ_{ph} .

The total probability for an electron to undergo a collision with a phonon is, obviously, defined by the sum of the probabilities $W = W_+ + W_-$, i.e. by the sum of the probabilities of collision with the emission and the absorption of a phonon. The quantities W_+ and W_- are given by formulae (56.4) and (56.5). For an actual calculation of W it is necessary to make the following simplifying assumptions concerning the quantities ω_f and $E(\mathbf{k})$ involved in these formulae:

(1) We shall assume that for all frequencies

$$\omega_f = cf, \quad (62.1)$$

where c is the velocity of sound, independent of f . This assumption cannot always be justified, particularly for high temperatures. However, it does not lead to any substantial error in the final result even in those cases where relation (62.1) is not accurately fulfilled.

(2) We shall assume that $E(|\mathbf{k}|)$ is a quadratic function of $|\mathbf{k}|$

$$E(|\mathbf{k}|) = \alpha k^2, \quad (62.2)$$

where $\alpha = \hbar^2/2m^*$. This assumption is valid for electrons moving in an almost filled or an almost empty band, i.e. for strongly bound and for nearly free electrons.

Let us first consider the case of temperatures which are high in comparison with the Debye temperature Θ_c . Then in formulae (56.4) and (56.5) one can set

$$n_f \cong \frac{kT}{\hbar\omega_f} = \frac{kT}{\hbar cf} \gg 1. \quad (62.3)$$

Indeed, the number of phonons is defined by the Bose distribution

$$n_f = [\exp(\hbar cf/kT) - 1]^{-1}.$$

For high temperatures the exponent may be expanded in a series. We then get formula (62.3). For the probability W we now obtain

$$W = \frac{2\pi}{\hbar} \frac{2g^2(kT)}{9NM_c^2} \int [\delta(E_{\mathbf{k}-\mathbf{f}} - E_{\mathbf{k}} + \hbar cf) + \delta(E_{\mathbf{k}+\mathbf{f}} - E_{\mathbf{k}} - \hbar cf)] \frac{V df}{(2\pi)^3}. \quad (62.4)$$

We shall carry out in detail the transformations with the delta-functions, which are of methodological interest.

Let us consider the first integral

$$I_1 = \int \delta(E_{\mathbf{k}-\mathbf{f}} - E_{\mathbf{k}} + \hbar cf) df. \quad (62.5)$$

In a metal at high temperatures, because of the degeneracy of states, actually only electrons in the spread-out region of the Fermi distribution are mobile. The energy of the electrons is $E_{\text{el}} \sim E_{\text{max}} \gg kT$, where E_{max} is the Fermi level; the wave numbers of these electrons are $k \sim k_{\text{max}} = \pi/a$.

All phonons up to those with wave number $f = f_{\text{max}} = \pi/a$, the latter being represented in the largest number, are excited in the lattice. The energy of the phonons is $\hbar\omega_f \leq \hbar(\omega_f)_{\text{max}} \sim kT$. Thus the wave vectors of the electrons and the phonons are quantities of the same order. On the contrary, the energy of the phonons is much smaller than that of the electrons. When an electron collides with a phonon the electron energy changes relatively little, while its wave vector changes considerably. This means that a considerably change in the direction of motion of the electron takes place in each collision. Inequality (62.3) is automatically fulfilled for all phonons; it is weaker than the inequality $f \leq f_{\text{max}}$.

Passing to the calculation of I_1 , we write

$$\begin{aligned} I_1 &= \int \delta\{\alpha(\mathbf{k} - \mathbf{f})^2 - \alpha k^2 + \hbar cf\} f^2 df \sin \theta d\theta d\varphi = \\ &= 2\pi \int \delta(-2\alpha k f \cos \theta + \alpha f^2 + \hbar cf) f^2 df \sin \theta d\theta. \end{aligned} \quad (62.5')$$

We carry out the integration over the angle by introducing the new variable $u = -2\alpha k f \cos \theta + \alpha f^2 + \hbar c f$. Then

$$I_1 = 2\pi \int f^2 df \frac{1}{2\pi\alpha k f} \int_{u_1}^{u_2} \delta(u) du,$$

where

$$u_1 = -2\alpha k f + \alpha f^2 + \hbar c f \quad (\text{which corresponds to } \theta = 0),$$

$$u_2 = 2\alpha k f + \alpha f^2 + \hbar c f \quad (\text{which corresponds to } \theta = \pi).$$

It is obvious that

$$\int_{u_1}^{u_2} \delta(u) du = \begin{cases} 0, & \text{if } u_1 \text{ and } u_2 \text{ have the same sign,} \\ 1, & \text{if } u_1 \text{ and } u_2 \text{ have different signs.} \end{cases}$$

The value of u_2 is essentially positive. Hence the value of the integral $I_1 \neq 0$, if

$$u_1 = -2\alpha k f + \alpha f^2 + \hbar c f < 0$$

or

$$f < 2k - (\hbar c / \alpha). \quad (62.6)$$

Since $k \lesssim k_{\max} = \pi/a$ is always true, inequality (62.6) does not impose any restrictions upon the integration over f . By virtue of this we obtain

$$I_1 = \frac{\pi}{2\alpha k} f_{\max}^2 = \frac{\pi m^* f_{\max}^2}{\hbar^2 k}.$$

Exactly the same calculation gives for the integral I_2

$$I_2 = \int \delta(E_{\mathbf{k}+\mathbf{f}} - E_{\mathbf{k}} - \hbar c f) df = I_1.$$

Substituting the values of I_1 and I_2 into the expression for W , we find

$$W = \frac{8\pi^2 g^2 (kT) V m^* f_{\max}^2}{9MN(2\pi)^3 \hbar^3 k c^2}.$$

We introduce the electron velocity $v_{\text{el}} = \hbar k / m^*$, substitute $f_{\max} = \pi/a$ for f_{\max} , and set the volume of the crystal equal to Na^3 . For the mean free path we then obtain the expression

$$\lambda_{\text{ph}} = \frac{v_{\text{el}}}{W} = \frac{9Mc^2 v_{\text{el}}^2 \hbar^2}{ng^2 (kT) a} \quad (62.7)$$

It is useful to estimate numerically the quantities involved in λ_{ph} .

The velocity of sound c is usually of the order of 2×10^5 cm/sec. The energy of an electron near the Fermi surface is $E_{\max} \sim p_{\max} v_{\text{el}} \sim (\hbar/a) v_{\text{el}} \simeq 2-3$ eV. Hence v_{el} is of the order of 10^8 cm/sec. The quantity g , defined by formula (55.12), is in order of magnitude the same as the kinetic energy of a nearly free electron. But the energy of nearly free electrons is close to the Fermi energy (since mobile electrons are in the spread-out region of the Fermi distribution), so that $g \sim 2-3$ eV. Finally, for metals of medium atomic weight $Mc^2 \sim 1$ eV. Consequently, in order of magnitude

$$\lambda_{\text{ph}} \sim \frac{3Mc^2}{g} \left(\frac{\hbar v_{\text{el}}}{a} \right) \frac{1}{g} \frac{\hbar v_{\text{el}}}{kT} \sim \frac{3\hbar v_{\text{el}}}{kT} \sim \frac{3\epsilon_{\max}}{kT} a. \quad (62.8)$$

For a Fermi energy $\epsilon_{\max} \sim 2$ eV and room temperature ($kT \simeq 0.025$ eV)

$$\lambda_{\text{ph}} \simeq 200-250 a.$$

This value of λ_{ph} is in good agreement with experimental data.

As is seen from (62.7), $\lambda_{\text{ph}} \sim 1/T$. Other quantities involved in (62.8) are in practice constant for a given metal.

Let us now consider the case of low temperatures $kT < \Theta_c$. Then the energy of excited phonons is $\hbar\omega \sim kT$, and their wave numbers are

$$f \sim \frac{\hbar\omega}{ch} \ll f_{\max} = \frac{\pi}{a}.$$

The energy of the electrons lies near the Fermi surface, i.e. is, as before, of the order of $E \sim E_{\max}$. The wave numbers of the electrons are $k \sim k_{\max} \sim \pi/a$.

Since $k \gg f$, the energy conservation law for the collision between an electron and a phonon can be written in the form

$$\frac{\hbar^2}{2m^*} (k + f)^2 - \frac{\hbar^2}{2m^*} k^2 - \hbar\omega_f \approx \frac{\hbar^2}{m^*} k \cdot f - \hbar\omega_f = 0.$$

Hence it follows that

$$\cos \theta = \frac{c}{(\hbar k/m^*)} \sim \frac{c}{v_{el}} \ll 1,$$

i.e. $\theta \sim \frac{1}{2}\pi$. This means that the phonon is emitted perpendicular to the direction of motion of the electron. The latter is deflected at a small angle ϑ — the angle between the vector \mathbf{k} and the vector $\mathbf{k} + \mathbf{f}$. In order of magnitude the angle ϑ is equal to

$$\vartheta \sim \frac{f}{k} \sim \frac{\hbar f}{\hbar k} \sim \frac{\hbar\omega}{\hbar c f_{\max}} \sim \frac{\hbar\omega}{\hbar\omega_{\max}} \sim \frac{T}{\Theta_c}.$$

The change of the electron wave vector in each collision is of the order of

$$\Delta k \sim k \cos \vartheta - k \sim \frac{1}{2}k\vartheta^2 \sim k \left(\frac{T}{\Theta_c} \right)^2.$$

Since each collision with a phonon leads to only a small change in the direction of motion, the electron must undergo a large number of collisions P to be scattered at a large angle. P is defined by the relation

$$P \cdot \Delta k \sim k$$

or

$$P \sim (\Theta_c/T)^2. \quad (62.9)$$

Let us now find the probability of scattering with a small change in the energy and momentum of an electron colliding with a phonon. The probability of collision with the absorption of a phonon is found very simply

$$W_- = \frac{4\pi g^2 V}{9MN} \int \frac{n_f f^2}{\omega_f} \delta(E_k + \hbar\omega_f - E_{k+f}) \frac{f^2 df \sin \theta d\theta d\varphi}{(2\pi)^3} =$$

$$\begin{aligned}
 &= \frac{8\pi^2 g^2 V}{9MNc} \int \frac{n_f f^3 df}{(2\pi)^3} \delta \left(-\frac{\hbar^2}{m^*} kf \cos \theta + \hbar \omega_f \right) \sin \theta d\theta = \\
 &= \frac{1}{\pi} \frac{g^2 V m^*}{9MNc \hbar^2 k} \int_0^{f_{\max}} \frac{f^2 df}{e^{\hbar cf/kT} - 1} = \frac{1}{\pi (\hbar c)^3} \frac{g^2 V m^* (kT)^3}{9MNc \hbar^2 k} \int_0^{\hbar cf_{\max}/kT} \frac{z^2 dz}{e^z - 1}.
 \end{aligned}$$

Since

$$\frac{\hbar cf_{\max}}{kT} = \frac{\hbar \omega_{\max}}{kT} = \frac{\Theta_c}{T} \gg 1,$$

the upper limit of the integral can be replaced by infinity. Then

$$W_- = \frac{1}{9\pi} \frac{g^2 a^3 m^*}{M c \hbar^2 k} \left(\frac{kT}{\hbar c} \right)^3 \int_0^\infty \frac{z^2 dz}{e^z - 1}.$$

This last integral is calculated in Appendix IV (vol. 2):

$$\int_0^\infty \frac{z^2 dz}{e^z - 1} = \frac{1}{3} \pi^2.$$

Hence

$$\begin{aligned}
 W_- &\approx \frac{g^2 a^3 (kT)^3}{9 v_{el} M \hbar^4 c^4} \approx \frac{a^3}{9} \left(\frac{g}{Mc^2} \right) \frac{g}{\hbar v_{el}} \left(\frac{kT}{\hbar c f_{\max}} \right)^3 f_{\max}^3 c \approx \\
 &\approx \frac{a^3}{9} \left(\frac{g}{Mc^2} \right) \left(\frac{ga}{\hbar v_{el}} \right) \left(\frac{kT}{\hbar c f_{\max}} \right) \frac{c}{a} \frac{\pi^3}{a^3} \approx 3 \frac{c}{a} \left(\frac{T}{\Theta_c} \right)^3.
 \end{aligned}$$

In calculating the probability of transition with the emission of a phonon it is necessary to take into account the fact that for the electron a transition into an occupied state is forbidden. This means that the electron cannot emit phonons with an energy exceeding the width of the spread-out region of the Fermi distribution.

If $\hbar \omega_f = \hbar cf > kT$, then, by emitting the corresponding phonon, the electron would be obliged to pass into an occupied state. Hence only the emission of phonons with wave numbers $f \leq kT/\hbar c$ is possible. Taking this fact into

account, we have

$$W_+ \approx \frac{4\pi g^2 V}{9MN(2\pi)^3 c} \int \frac{f^3 df \sin \theta d\theta d\varphi}{e^{\hbar c f / kT} - 1} \delta \left(\frac{\hbar^2 k f \cos \theta}{m^*} - \hbar c f \right) \approx W_- . \quad (62.10)$$

The mean free path of an electron before the first collision is

$$\lambda_{ph}^{(1)} = \frac{v_{el}}{W} \sim a \left(\frac{v_{el}}{c} \right) \left(\frac{\Theta_c}{T} \right)^3 . \quad (62.11)$$

Besides the mean free path between two collisions, one can introduce the mean transport free path λ_{tr} which takes into account the effectiveness of the collisions. The corresponding cross section is defined as

$$\sigma_{tr} = \int \sigma (1 - \cos \theta) d\Omega \sim \int \frac{1}{2} \sigma \theta^2 d\Omega .$$

The transport mean free path λ_{tr} , i.e. the mean free path before an effective change of momentum, is defined in classical kinetics as $\lambda_{tr} = 1/N\sigma_{tr}$. In our approximation we introduce the quantity λ_{tr} , defining it by the relation

$$\lambda_{tr} \approx \frac{v_{el}}{W} P \approx \frac{av_{el}}{c} \left(\frac{\Theta_c}{T} \right)^5 . \quad (62.12)$$

It increases very rapidly, as $1/T^5$, with decreasing temperature. At very low temperatures, of the order of a few degrees, λ_{tr} is very large compared to a and reaches macroscopic values.

Knowing the mean free path, one can find the electrical conductivity by formula (58.8).

The electrical conductivity of metals σ is expressed by the formulae

$$\sigma = \frac{e^2 \lambda_{ph} n}{mv} = \frac{9ne^2 M c^2 \hbar^2 v_{ph}}{\pi m g^2 (kT) a} \text{ for high temperatures ,} \quad (62.13)$$

$$\sigma = A \frac{ne^2 \hbar M c^2 v_{el}}{m g^2 c} \left(\frac{\Theta_c}{T} \right)^5 \text{ for low temperatures ,} \quad (62.14)$$

where A is a numerical coefficient.

All quantities involved in (62.13) and (62.14) are known and can be found from independent measurements. Agreement between theory and experiment turns out to be good. This is somewhat unexpected, taking into account the

large number of simplifications made in the calculations and the fact that the effect of collisions between electrons was neglected. It should be noted that for $T \rightarrow 0$ the electrical conductivity of real (non-superconducting) metals does not increase up to infinitely large values but tends to a constant limit independent of temperature. This is the so-called residual resistance, due to the scattering of electrons by impurities and non-uniformities. Its value is determined by the concentration of impurities and is obtained by substituting λ_{imp} for λ_{ph} in the formula for σ .

Knowing the mean free path, one can find the electronic thermal conductivity of a metal. For high temperatures it is expressed by the formula

$$\kappa_{\text{el}} \approx v_{\text{el}} \lambda_{\text{ph}} C_V^{\text{el}} \approx \lambda_{\text{ph}} \frac{k T v_{\text{el}}}{\epsilon_{\text{max}}} n \sim \text{const} \quad (62.15)$$

and does not depend on the temperature.

It is obvious that for $T \gg \Theta_c$ the so-called Wiedemann–Franz law $T\sigma/\kappa = \text{const}$ holds. For low temperatures the thermal diffusivity is determined by the mean free path λ_{ph} . Indeed, for a single collision with a phonon the energy of an electron changes by an amount of the order of the phonon energy, i.e. of the order of kT . This change of energy corresponds to the possible amount of energy transport by electrons. The transport of large energies is impossible since the electron cannot give up an energy considerably exceeding kT . Hence the mechanism of energy transport by electrons amounts to transfers of energies of the order of kT in each collision with a phonon, which corresponds to deflections of the electron at small angles.

The thermal conductivity of electrons for $kT \ll \Theta_c$ is expressed by the formula

$$\kappa_{\text{el}} \approx \lambda_{\text{ph}}^{(1)} C_V^{\text{el}} v_{\text{el}} \sim \lambda_{\text{ph}}^{(1)} \frac{kT}{\epsilon_{\text{max}}} n v_{\text{el}} \sim \frac{1}{T^2}, \quad (62.16)$$

where $\lambda_{\text{ph}}^{(1)}$ is the mean free path before the first collision, which is given by (62.11)

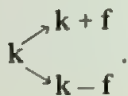
It is clear that in this case the Wiedemann–Franz law does not hold. The ratio of electrical conductivity to thermal conductivity turns out to depend on the temperature. We cannot dwell in this book on other phenomena associated with the electron gas in metals, and refer the reader to the specialist literature*.

*See footnote on p. 326.

§63. The collision integral for electrons in a metal

We now pass to a logical treatment of the kinetic equation and to the proof of the existence of a mean free path for electrons in a metal. For this we must first of all obtain the expression for the collision integral. We shall take into account only collisions between electrons and phonons.

Let us consider a certain state \mathbf{k} of an electron. The electron may leave the given state \mathbf{k} in two ways: by absorbing a phonon \mathbf{f} or by emitting a phonon \mathbf{f} , i.e.



The number of electrons leaving the state \mathbf{k} in unit time by emitting a phonon is equal to

$$\Phi(|\mathbf{k}|)(1 - \Phi(|\mathbf{k} - \mathbf{f}|)) dW_+(\mathbf{k} \rightarrow \mathbf{k} - \mathbf{f}),$$

where $dW_+(\mathbf{k} \rightarrow \mathbf{k} - \mathbf{f})$ is given by formula (56.5), and the factor $(1 - \Phi(|\mathbf{k} - \mathbf{f}|))$ is introduced to take into account the Pauli principle. Because of this factor no electrons pass into the occupied state for which $\Phi(\mathbf{k} - \mathbf{f}) = 1$.

The number of electrons leaving the state \mathbf{k} by absorbing a phonon is equal to

$$\Phi(\mathbf{k})(1 - \Phi(\mathbf{k} + \mathbf{f})) dW_-(\mathbf{k} \rightarrow \mathbf{k} + \mathbf{f}),$$

where $dW_-(\mathbf{k} \rightarrow \mathbf{k} + \mathbf{f})$ is defined by (56.4). The total number of electrons leaving the state \mathbf{k} per second is obtained by integrating the sum of the above expressions over all possible values of \mathbf{f} , i.e.

$$I_- = \int \{ \Phi(\mathbf{k})(1 - \Phi(\mathbf{k} - \mathbf{f})) dW_+(\mathbf{k} \rightarrow \mathbf{k} - \mathbf{f}) + \Phi(\mathbf{k})(1 - \Phi(\mathbf{k} + \mathbf{f})) dW_-(\mathbf{k} \rightarrow \mathbf{k} + \mathbf{f}) \}.$$

* See, for example, R. Peierls, *Quantum theory of solids* (Clarendon Press, Oxford, 1966); A.H. Wilson, *The theory of metals* (Cambridge University Press, 1953); A. Sommerfeld and H. Bethe, *Handbuch der Physik*, Vol. 24/2 (Springer, Berlin, 1933); C. Kittel, *Quantum theory of solids* (Wiley, New York, London, 1969); J.M. Ziman, *Principles of the theory of solids* (Cambridge University Press, Cambridge, 1964).

The integration is carried out over all values of $|\mathbf{f}|$ and over all orientations of the vector \mathbf{f} .

Let us write down the number of electrons arriving per unit time into the state \mathbf{k} as a result of the emission and absorption of phonons by electrons in other states. Reasoning analogous to the above gives

$$I_+ = \int \{ \Phi(\mathbf{k} + \mathbf{f}) (1 - \Phi(\mathbf{k})) dW_+(\mathbf{k} + \mathbf{f} \rightarrow \mathbf{k}) + \\ + \Phi(\mathbf{k} - \mathbf{f}) (1 - \Phi(\mathbf{k})) dW_-(\mathbf{k} - \mathbf{f} \rightarrow \mathbf{k}) \}.$$

Setting up the balance of electrons, we can write

$$I_{\text{coll}} = I_+ - I_-.$$

Substituting the values of I_+ and I_- and of the transition probabilities dW_+ and dW_- involved in them, according to the formulae (56.4) and (56.5) we can write the integral I_{coll} in the form

$$I_{\text{coll}} = \frac{4\pi}{9\hbar} \left(\frac{\hbar}{MN} \right) g^2 V \int \frac{f^2}{\omega_f} \{ \{ (n_f + 1) \Phi(\mathbf{k} + \mathbf{f}) (1 - \Phi(\mathbf{k})) \delta(E_{\mathbf{k}+\mathbf{f}} - \\ - E_{\mathbf{k}} - \hbar\omega_f) + n_f (1 - \Phi(\mathbf{k})) \Phi(\mathbf{k} - \mathbf{f}) \delta(E_{\mathbf{k}-\mathbf{f}} - E_{\mathbf{k}} + \hbar\omega_f) \} - \\ - \{ (n_f + 1) \Phi(\mathbf{k}) (1 - \Phi(\mathbf{k} - \mathbf{f})) \delta(E_{\mathbf{k}} - E_{\mathbf{k}-\mathbf{f}} - \hbar\omega_f) + \\ + n_f \Phi(\mathbf{k}) (1 - \Phi(\mathbf{k} + \mathbf{f})) \delta(E_{\mathbf{k}} - E_{\mathbf{k}+\mathbf{f}} + \hbar\omega_f) \} \} \frac{d\mathbf{f}}{(2\pi)^3}.$$

Combining processes with the same delta-functions and substituting I_{coll} into the kinetic equation, we obtain

$$\frac{\partial \Phi}{\partial t} + \mathbf{v} \cdot \frac{\partial \Phi}{\partial \mathbf{r}} + \frac{\mathbf{F}}{\hbar} \cdot \frac{\partial \Phi}{\partial \mathbf{k}} = \frac{4\pi V}{9\hbar} \left(\frac{\hbar}{MN} \right) g^2 \int \frac{f^2}{\omega_f} \times \\ \times \{ [\Phi(\mathbf{k} + \mathbf{f}) (1 - \Phi(\mathbf{k})) (n_f + 1) - \Phi(\mathbf{k}) (1 - \Phi(\mathbf{k} + \mathbf{f})) n_f] \times \\ \times \delta(E_{\mathbf{k}+\mathbf{f}} - E_{\mathbf{k}} - \hbar\omega_f) + [n_f (1 - \Phi(\mathbf{k})) \Phi(\mathbf{k} - \mathbf{f}) -$$

$$-(n_f + 1) \Phi(\mathbf{k}) (1 - \Phi(\mathbf{k} - \mathbf{f})) [\delta(E_{\mathbf{k}-\mathbf{f}} - E_{\mathbf{k}} + \hbar\omega_f)] \frac{d\mathbf{f}}{(2\pi)^3}. \quad (63.1)$$

Here in combining the terms we have made use of the even nature of the delta-function

$$\delta(E_{\mathbf{k}+\mathbf{f}} - E_{\mathbf{k}} - \hbar\omega_f) \equiv \delta(E_{\mathbf{k}} - E_{\mathbf{k}+\mathbf{f}} + \hbar\omega_f).$$

The kinetic equation represents the quantum analogue of eq. (57.8). Comparison with the classical equation (57.8) shows that: (1) the transition probability containing the unknown cross section is replaced by known relations for dW_+ and dW_- ; (2) the Pauli principle is taken into account by introducing factors of the type $(1 - \Phi(\mathbf{k}))$ into the number of transitions, while the factors n_f and $(1 + n_f)$ take into account the statistical weights of states of the system of phonons.

Let us first of all convince ourselves of the fact that in an equilibrium state the distribution function of electrons $\Phi(\mathbf{k}) \equiv \Phi_0$ represents a Fermi-Dirac distribution, and that the distribution function of phonons $n_f \equiv n_0$ represents a Bose-Einstein distribution. In an equilibrium state $\partial\Phi/\partial t = 0$, $\mathbf{v} = 0$, $\mathbf{F} = 0$, so that

$$I_{\text{coll}} = 0. \quad (63.2)$$

It is easily seen that condition (63.2) is fulfilled if each of the brackets in the integrand of (63.1) reduces to zero, i.e.

$$\Phi_0(\mathbf{k} + \mathbf{f})(1 - \Phi_0(\mathbf{k}))(n_f + 1) - \Phi_0(\mathbf{k})(1 - \Phi_0(\mathbf{k} + \mathbf{f}))n_f = 0, \quad (63.3)$$

$$E(\mathbf{k} + \mathbf{f}) = E(\mathbf{k}) + \hbar\omega_f,$$

$$\Phi_0(\mathbf{k} - \mathbf{f})(1 - \Phi_0(\mathbf{k}))n_f - \Phi_0(\mathbf{k})(1 - \Phi_0(\mathbf{k} - \mathbf{f}))(n_f + 1) = 0, \quad (63.4)$$

$$E(\mathbf{k} - \mathbf{f}) = E(\mathbf{k}) - \hbar\omega_f.$$

The solution of the functional equations (63.3) and (63.4) with the subsidiary relations expressing energy conservation is carried out according to the usual scheme. One then obtains

$$\Phi_0 = \frac{1}{e^{(\epsilon - \mu)/kT} + 1}, \quad n_f = \frac{1}{e^{\hbar\omega/kT} - 1}.$$

Knowing the equilibrium distribution functions of electrons and phonons and assuming the departure from the equilibrium state to be small, one can solve the integro-differential equation (63.1).

§64. A solution of the kinetic equation

In what follows we shall confine ourselves to the calculation of the electrical conductivity of a metal for high temperatures $T \gg \Theta$, where Θ is the Debye temperature. In a time independent uniform electric field the kinetic equation can be written as

$$\frac{e\mathcal{E}}{\hbar} \cdot \frac{\partial \Phi}{\partial \mathbf{k}} = I_{\text{coll}}. \quad (64.1)$$

We shall seek its solution in the form

$$\Phi = \Phi_0 + \Phi', \quad (64.2)$$

where the correction to the distribution function is equal to

$$\Phi'(\mathbf{k}) = \Phi_1(|\mathbf{k}|) k_z = \Phi_1(|\mathbf{k}|) k \cos \vartheta, \quad (64.3)$$

ϑ is the angle between the directions of the vector \mathbf{k} and of the field \mathcal{E} .

In calculating the integral it is necessary to make certain assumptions about the form of the functions $E(k)$ and ω_f . We shall assume that $\omega_f = cf$, where c is the velocity of sound.

Further, we shall assume that the electron energy is of the form

$$E = \alpha k^2, \quad (64.4)$$

as in the case of nearly free electrons and strongly bound electrons near the boundary of a zone.

We transform the collision integral under these simplifying assumptions, substituting into it ω as defined by formula (47.16). In carrying out the integration we introduce polar coordinates with the polar axis directed along the vector \mathbf{k} .

Let us write the collision integral. In order to abbreviate the formulae we shall confine ourselves to its first part which contains the delta-function corresponding to the emission process

$$I_{\text{coll}} = \frac{4\kappa}{9\hbar(2\pi)^3} \left(\frac{\hbar}{MN} \right) g^2 V (I_1 + I_2),$$

$$I_1 = \int \frac{f^2}{\omega_f} \{ [\Phi_0(\mathbf{k} + \mathbf{f}) + \Phi'(\mathbf{k} + \mathbf{f})] [1 - \Phi_0(\mathbf{k}) - \Phi'(\mathbf{k})] (n_f + 1) -$$

$$- [\Phi_0(\mathbf{k}) + \Phi'(\mathbf{k})] [1 - \Phi_0(\mathbf{k} + \mathbf{f}) - \Phi'(\mathbf{k} + \mathbf{f})] n_f \} \times$$

$$\times \delta(E_{\mathbf{k}+\mathbf{f}} - E_{\mathbf{k}} - \hbar\omega_f) f^2 df \sin \theta d\theta d\varphi. \quad (64.5)$$

Further we shall make use of the equilibrium conditions (63.3) and (63.4), and shall disregard the products of the functions $\Phi'(\mathbf{k} + \mathbf{f}) \Phi'(\mathbf{k})$. We shall substitute into the delta-function the expression (64.4) for the energy

$$I_1 = \int \frac{f^3}{c} \{ \Phi'(\mathbf{k} + \mathbf{f}) [1 - \Phi_0(\mathbf{k})] (n_f + 1) - \Phi'(\mathbf{k}) \Phi_0(\mathbf{k} + \mathbf{f}) (n_f + 1) -$$

$$- \Phi'(\mathbf{k}) [1 - \Phi_0(\mathbf{k} + \mathbf{f})] n_f + \Phi'(\mathbf{k} + \mathbf{f}) \Phi_0(\mathbf{k}) n_f \} \times$$

$$\times \delta(2\alpha k f \cos \theta + \alpha f^2 - \hbar c f) \sin \theta d\theta d\varphi df.$$

In what follows we shall restrict ourselves to the case of high temperatures.

For high temperatures $n_f \gg 1$. Under this assumption one can make the substitution

$$n_f + 1 \simeq n_f,$$

$$n_f \simeq kT/\hbar c f.$$

Let us now write the expansion (64.2) in more detail:

$$\Phi(\mathbf{k}) = \Phi_0(E_{\mathbf{k}}) + k_{\mathcal{E}} \Phi_1(E_{\mathbf{k}}) = \Phi_0(E_{\mathbf{k}}) + k \cos \vartheta \Phi_1(E_{\mathbf{k}}),$$

$$\Phi(\mathbf{k} + \mathbf{f}) = \Phi_0(E_{\mathbf{k}+\mathbf{f}}) + k_{\mathcal{E}} \Phi_1(E_{\mathbf{k}+\mathbf{f}}) + f_{\mathcal{E}} \Phi_1(E_{\mathbf{k}+\mathbf{f}}) =$$

$$= \Phi_0(E_{\mathbf{k}} + \hbar\omega_f) + k \cos \vartheta \Phi_1(E_{\mathbf{k}} + \hbar\omega_f) + f \cos \gamma \Phi_1(E_{\mathbf{k}} + \hbar\omega_f).$$

Here $k_{\mathcal{E}}$ and $f_{\mathcal{E}}$ are the projections of the vectors \mathbf{k} and \mathbf{f} onto the direction of the electric field \mathcal{E} . We then find

$$I_1 = \int \frac{f^2 kT}{\hbar c^2} \{ (k \cos \vartheta + f \cos \gamma) \Phi_1(E_{\mathbf{k}+\mathbf{f}}) - \\ - k \cos \vartheta \Phi_1(E_{\mathbf{k}}) \} \delta(2\alpha k f \cos \theta + \alpha f^2 - \hbar c f) \sin \theta \, d\theta \, d\varphi \, df.$$

Since $E_{\mathbf{k}+\mathbf{f}} = E_{\mathbf{k}} + \hbar \omega_f$ and the electron energy is of the order of ϵ_{\max} , we can set

$$\Phi_1(E_{\mathbf{k}+\mathbf{f}}) \simeq \Phi_1(E_{\mathbf{k}}).$$

Hence

$$I_1 = \frac{kT}{\hbar c^2} \Phi_1(E_{\mathbf{k}}) \int f^3 \cos \gamma \delta(2\alpha k f \cos \theta + \alpha f^2 - \hbar c f) \sin \theta \, d\theta \, d\varphi \, df.$$

To carry out the integration over angles, we express $\cos \gamma$ in terms of $\cos \theta$ and $\cos \vartheta$

$$\cos \gamma = \cos \vartheta \cos \theta + \sin \vartheta \sin \theta \cos \varphi.$$

Then

$$I_1 = \frac{kT}{\hbar c^2} \Phi_1(E_{\mathbf{k}}) \int f^3 (\cos \theta \cos \vartheta + \sin \vartheta \sin \theta \sin \varphi) \times \\ \times \delta(2\alpha k f \cos \theta + \alpha f^2 - \hbar c f) \sin \theta \, d\theta \, d\varphi \, df = \\ = \frac{2\pi kT}{\hbar c^2} \Phi_1(E_{\mathbf{k}}) \cos \vartheta \int_0^\pi f^3 \cos \theta \delta(2\alpha k f \cos \theta + \alpha f^2 - \hbar c f) \sin \theta \, d\theta \, df.$$

In integrating over φ the term proportional to $\sin \varphi$ reduces to zero. The integration over the angle θ can be carried out by introducing the new variable

$$2\alpha k f \cos \theta + \alpha f^2 - \hbar c f = u.$$

Under this transformation we obtain

$$I_1 = \frac{2\pi(kT)}{\hbar c^2} \frac{\cos \vartheta}{2\alpha k} \Phi_1(E_{\mathbf{k}}) \int f^2 df \int_{u_1}^{u_2} \left(\frac{u}{2\alpha k f} + \frac{\hbar c f}{2\alpha k f} - \frac{\alpha f^2}{2\alpha k f} \right) \delta(u) \, du.$$

As we have already seen in §62,

$$\int_{u_1}^{u_2} \delta(u) du = \int_{-\infty}^{+\infty} \delta(u) du = 1.$$

Furthermore, since for all possible values of f , u_1 and u_2 have different signs

$$\int_{u_1}^{u_2} u \delta(u) du = \int_{-\infty}^{+\infty} u \delta(u) du = 0.$$

Taking these equalities into account, we obtain finally

$$\begin{aligned} I_1 &= \frac{\pi(kT)}{\hbar c^2} \frac{\cos \vartheta \Phi_1(E_k)}{\alpha k} \int_0^{f_{\max}} \left(\frac{\hbar c}{2\alpha k} - \frac{\alpha f}{2\alpha k} \right) f^2 df = \\ &= \frac{\pi(kT)}{\hbar c^2} \frac{\cos \vartheta}{2\alpha k} \left(\frac{\hbar c f_{\max}^3}{3\alpha k} - \frac{f_{\max}^4}{4k} \right) \Phi_1(E_k). \end{aligned}$$

Calculating the second part of the collision integral gives

$$\begin{aligned} I_2 &= \int \frac{f^2}{\omega_f} \{ [n_f(1 - \Phi(\mathbf{k})\Phi(\mathbf{k}-\mathbf{f}) - (n_f+1)\Phi(\mathbf{k}) \times \\ &\quad \times (1 - \Phi(\mathbf{k}-\mathbf{f}))] \} \delta(E_{\mathbf{k}-\mathbf{f}} - E_{\mathbf{k}} + \hbar\omega_f) d\mathbf{f} = \\ &= \frac{\pi(kT)}{\hbar c^2} \frac{\cos \vartheta}{2\alpha k} \left(-\frac{\hbar c^3}{3\alpha k} - \frac{f_{\max}^4}{4k} \right) \Phi_1(E_k). \end{aligned}$$

By virtue of (64.5) and the expressions for I_1 and I_2 , we find the expression for the collision integral

$$I_{\text{coll}} = - \frac{\pi^2(kT)g^2 f_{\max}^4}{9\hbar c^2(2\pi)^3 \alpha k^2} \left(\frac{V}{MN} \right) \cos \vartheta \Phi_1(E_k). \quad (64.6)$$

We now transform the left-hand side of the kinetic equation. Assuming the force \mathbf{F} acting on the electron to be weak, we have

$$\mathbf{F} \cdot \frac{\partial \Phi_0}{\partial \mathbf{p}} \simeq \frac{e\mathbf{E}}{\hbar} \frac{\partial \Phi_0}{\partial E_{\mathbf{k}}} \frac{\partial E_{\mathbf{k}}}{\partial \mathbf{k}} \cdot \frac{\partial \mathbf{k}}{\partial k_z} = \frac{e\mathbf{E} \cos \vartheta}{\hbar} 2\alpha k \frac{\partial \Phi_0}{\partial E_{\mathbf{k}}}.$$

Equating this last expression and I_{coll} , we obtain the value of the correction to the equilibrium distribution function

$$\Phi_1 = -e\mathbf{E} \frac{\partial \Phi_0}{\partial E_{\mathbf{k}}} \left(\frac{18\alpha^2 k^3 (Mc^2)}{\pi^2 k T g^2} \left(\frac{N}{V} \right) \frac{1}{f_{\text{max}}^4} \right) (2\pi)^3. \quad (64.7)$$

Comparison of (64.7) with (27.11) shows that

$$\lambda_{\text{ph}} = \frac{144\alpha^2 k_0^4 (Mc^2) \pi N}{k T g^2 V f_{\text{max}}^4}. \quad (64.8)$$

In accordance with (80.13) of Part III the value of the electron energy is assumed to be equal to the Fermi energy $\alpha k_0^2 = \mu \simeq E_{\text{max}}$, where μ is the partial potential of the equilibrium distribution of the electrons. We shall further show that the result obtained for λ_{ph} is in agreement with formula (62.8). For this we write the expression for λ_{ph} in the form

$$\lambda_{\text{ph}} = 144 \frac{E_{\text{el}}^2}{g} \left(\frac{Mc^2}{g} \right) \frac{a}{\pi^3 k T} \simeq 5 \left(\frac{Mc^2}{g} \right) \frac{E_{\text{el}}}{g} \frac{E_{\text{el}}}{k T} a \simeq 5 \frac{E_{\text{el}}}{k T} a.$$

This value of λ_{ph} is in very good agreement with the qualitative results obtained in §62.

The solution of the kinetic equation for the case of temperatures lower than the Debye temperature can be found in an analogous way. Then the same expression as (62.12) is obtained for λ_{tr} . We shall not dwell on these rather cumbersome calculations. The theory is in good agreement with experiment for metals having a crystal lattice with a high symmetry.

Effects of anisotropy turn out to be very important in metals having a lattice with low symmetry. The difference between the tensor components σ_{ik} in different crystallographic directions sometimes turns out to be considerable. To construct a theory of kinetic phenomena taking into account anisotropy it is necessary, first of all, to study the character of the Fermi surface in more detail.

In the case of metals with a relatively small amount of impurity, or in the case of alloys with a small concentration of one of the components or

ordered solid solutions, the approximations of the theory turn out to be valid. However, in the case of strongly alloyed metals, disordered solid solutions and liquid metals, the basic assumption of the theory, the existence of a regular periodic structure, is not fulfilled.

As we have pointed out in §39, the electrical conductivity of such systems is calculated by means of the mathematical apparatus of time correlation functions.

§65. Superconductivity

In the preceding sections we have considered certain features of metallic conductors. Our considerations would be incomplete if we did not discuss, at least briefly, the phenomenon of superconductivity. The macroscopic theory of superconductivity has already been discussed (§21 of Part IV).

There has recently been success in constructing an adequate microscopic theory of superconductivity, which we shall outline here.

For some time it has been assumed that there is an analogy between the phenomena of superfluidity and superconductivity. That is the current in a superconductor, which is not weakened by ohmic resistance, has naturally been thought of as a superfluid current of electrons in a lattice.

As was explained in §5, superfluidity arises in a system of particles if the energy spectrum of its collective excitations satisfies certain requirements. These requirements are not related directly to the statistics of the particles constituting the system. However, until recently the spectrum of collective excitations satisfying the condition of superfluidity could be obtained only for an imperfect Bose gas. Qualitatively the reason for this can be understood by the following reasoning which allows the difference between the Fermi and Bose systems to be elucidated.

The particles of a Bose gas in the superconducting state form a condensate accumulating in a state with momentum equal to zero. Repulsive forces between the particles ensure the appearance of collective motion in the system. The excitations have an energy spectrum satisfying the condition of superfluidity (5.23). To small impulses of excitation of the system there corresponds a small energy.

In the case of a system of electrons, an ideal gas of fermions, the situation is essentially different. In such a system the condensation of particles in momentum space is impossible. The particles successively occupy lower quantum states down to the Fermi level. The appearance of a very small excitation in such a system means that one of the particles leaves a state at the Fermi sur-

face and goes over into an unoccupied (excited) state. Two unpaired 'particles' appear in the system: an electron in an unoccupied state and a hole, with momenta close to the momentum of particles at the Fermi surface, i.e. with momenta having a very large absolute value.

Thus in a system of fermions the condition of superfluidity $|v| < \epsilon/p$ for small ϵ and large p is not fulfilled if the value of the velocity is appreciable. The electrostatic interaction between fermions cannot change this situation. As has been said, condensation is impossible in a system of interacting fermions.

Electrons in a metal form a gas of fermions interacting with each other according to the Coulomb law, i.e. they undergo mutual repulsion. It seemed incomprehensible that a system of electrons could move in a metal without interacting with the crystal lattice of the metal.

One of the important stages in understanding the nature of superconductivity was the discovery of the isotope effect (see §21 of Part IV). From the existence of the isotope effect it followed that the interaction of electrons with lattice vibrations plays an important role in the phenomenon of superconductivity.

In this connection it is interesting to note that superconductivity does not arise in systems possessing good conductivity, for example in metals of the first group of the Mendeleev periodic system. On the contrary, metals which at ordinary temperatures have a considerably larger resistance, possess the property of superconductivity. Thus the phenomenon of superconductivity is observed in metals with a relatively strong interaction between the electron gas and phonon gas.

We have already mentioned in §56 that the absorption and emission of virtual phonons by electrons leads to a certain effective interaction between the electrons*. In 1956 L.N.Cooper published a note in which he pointed out that owing to the existence of a weak interaction (attraction) between electrons they may form certain bound states; electron pairs. These pairs possess an integer spin and, roughly speaking, the set of them can be considered as a Bose gas, which at low temperatures may possess the property of superfluidity. Proceeding from the concept of pairs, J.Bardeen, L.N.Cooper and J.R.Schrieffer developed a theory of the phenomenon of superconductivity**. In its most strict form the theory of this phenomenon was given by

* H.Fröhlich, Phys. Rev. 79 (1950) 845; J.Bardeen, Phys. Rev. 80 (1950) 567; L.N.Cooper, Phys. Rev. 104 (1956) 1180.

** J.Bardeen, L.N.Cooper and J.R.Schrieffer, Phys. Rev. 106 (1957) 162.

Bogoliubov*. We shall first present the qualitative picture of the formation of electron pairs, following the obvious derivation of Cooper.

Let us consider the motion of two electrons. We write the wave function of the system in the form

$$\psi(\mathbf{R}, \mathbf{r}) = \frac{1}{V^{\frac{1}{2}}} e^{i\mathbf{K} \cdot \mathbf{r}} \psi_0(\mathbf{r}, \mathbf{k}), \quad (65.1)$$

where in correspondence with (14.14) of Part V the wave function of the system is equal to the product of wave functions characterizing the relative motion and the motion of the center of mass. Here $\mathbf{R} = \frac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)$, $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, and \mathbf{K} is the wave vector of the system as a whole.

We write the wave function of the relative motion in the momentum representation

$$\psi_0 = \sum_{k > k_0} a_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{r}}. \quad (65.2)$$

We normalize the plane waves to a volume V and impose conditions of periodicity upon them. Because electron states with an energy lower than the Fermi energy are occupied, the summation in (65.2) is correspondingly bounded from below. Of course, such a bound from below on the summation is not a strictly consequential situation. In fact, it is necessary to consider the many-electron problem. Cooper reduces this problem to that of two interacting electrons on the background of the occupied Fermi sphere. The background electrons are taken into account in the form of a restriction on the summation, i.e. in expression (65.2) one sets $a_{\mathbf{k}} = 0$ for $k < k_0$.

The Schrödinger equation for two particles in the momentum representation has the form (see (48.12') of Part V)

$$(\epsilon_{\mathbf{K}} + \epsilon_{\mathbf{k}} - E) a_{\mathbf{k}} + \sum_{\mathbf{k}'} a_{\mathbf{k}'} \langle \mathbf{k} | H' | \mathbf{k}' \rangle = 0, \quad (65.3)$$

where

$$\langle \mathbf{k} | H' | \mathbf{k}' \rangle = \frac{1}{V} \int d\mathbf{r} e^{-i\mathbf{k} \cdot \mathbf{r}} H' e^{i\mathbf{k}' \cdot \mathbf{r}}$$

* N.N.Bogoliubov, V.V.Tolmachev and D.V.Shirkov, Fortschr. der Physik 6 (1958) 605.

and

$$\varepsilon_{\mathbf{K}} = \hbar^2 K^2 / 4m, \quad \epsilon_{\mathbf{k}} = \hbar^2 k^2 / m. \quad (65.4)$$

As has already been mentioned, the absorption and emission of phonons by electrons leads to a certain effective interaction between them. This interaction is analogous to the Coulomb interaction in electrodynamics due to photon exchange.

The exchange interaction might be obtained by means of Fröhlich's Hamiltonian (55.11). However, the derivation based on Fröhlich's Hamiltonian is complicated. Hence we shall consider a simplified Hamiltonian making it possible to correctly draw basic qualitative conclusions and requiring no complex calculations.

We assume that the energies of interacting electrons in the pair lie in the interval $\Delta\epsilon = \epsilon_m - \epsilon_0 = \hbar^2(k_m^2 - k_0^2)/2m$ above the Fermi surface ϵ_{\max} , so that we set

$$\begin{aligned} \langle k | H' | k' \rangle &= -F \quad \text{if } k_0 < k, k' < k_m \\ \langle k | H' | k' \rangle &= 0 \quad \text{if } k \text{ or } k' \text{ lies outside} \\ &\quad \text{the region mentioned.} \end{aligned} \quad (65.5)$$

Here F is a constant, and $(\hbar^2/m)(k_m^2 - k_0^2) \approx 2\hbar\omega \approx 0.2 \text{ eV}$ ($\hbar\omega$ is of the order of a certain effective phonon energy, and $\epsilon_0 = \hbar^2 k_0^2 / m$ is the energy of an electron at the Fermi surface).

We substitute expression (65.5) into eq. (65.3). We then have

$$a_{\mathbf{k}} = \frac{F \sum_{\mathbf{k}'} a_{\mathbf{k}'}}{\varepsilon_{\mathbf{K}} + \epsilon_{\mathbf{k}} - E}, \quad k_0 < k < k_m \quad (65.6)$$

and $a_{\mathbf{k}} = 0$ for k lying outside the interval mentioned. Further, we calculate the sum $\sum_{\mathbf{k}} a_{\mathbf{k}}$. It is equal to

$$\sum_{\mathbf{k}} a_{\mathbf{k}} = F \sum_{\mathbf{k}} \frac{\sum_{\mathbf{k}'} a_{\mathbf{k}'}}{\varepsilon_{\mathbf{K}} + \epsilon_{\mathbf{k}} - E}, \quad (65.7)$$

Hence we obtain

$$1 = -F \int_{\epsilon_0}^{\epsilon_m} \frac{N(K, \epsilon) d\epsilon}{E - \epsilon_K - \epsilon}. \quad (65.8)$$

Here, the energies ϵ_0 and ϵ_m correspond to the momenta $\hbar k_0$ and $\hbar k_m$, and we have passed from a summation to integration over the corresponding energy interval. $N(K, \epsilon)$ denotes the density of two-electron states with energy ϵ and total momentum K . Eq. (65.8) defines the energy of the system E . Owing to the narrowness of the energy interval in which the interaction takes place, the density of states $N(K, \epsilon)$ can be replaced by $N(K, \epsilon_0)$. After this the integration may be carried out in an elementary way. Solving the equation obtained for E , we have $E = E_0 = \epsilon_K + \epsilon_0 - \Delta$, where

$$\Delta = \frac{\epsilon_m - \epsilon_0}{e^{1/\beta} - 1} \quad (65.9)$$

and

$$\beta = N(K, \epsilon_0) F. \quad (65.10)$$

It is obvious that the state found with energy $E = E_0$ corresponds to a bound state. Indeed, eq. (65.8) has a solution only for a definite value of E . It can be seen that the continuous spectrum corresponding to the disintegration of the pair is separated from the bound state energy level by a gap of width Δ . The quantity Δ depends strongly on the density of states $N(K, \epsilon_0)$. The function N in its turn depends strongly on the total momentum K of the pair. To illustrate this dependence, let us examine fig. VI.13. The centres of the Fermi spheres of radius k_0 of the two electrons are separated by the value of the total momentum K of the system. The value of δk corresponds to that region in momentum space in which the interaction takes place. Since the total mo-

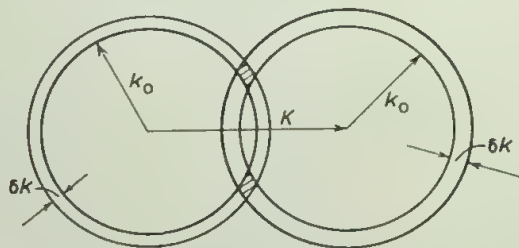


Fig. VI.13

mentum of the electrons equals K and each of the momenta of the individual electrons must lie in the region δk , it is obvious that the volume of the hatched region is proportional to the number of states $N(K, \epsilon_0)$. Hence it is seen that $N(K, \epsilon_0)$ has a sharp peak for a total momentum equal to zero. In this case the spheres are matched and the volume of the hatched region becomes a maximum. This corresponds to the fact that electrons with opposite momenta are actually bound into pairs. Thus the bound state of a system of two electrons arises even when there is a very weak interaction. We note that, in general, two microparticles can form bound states only when there is a sufficiently strong interaction. In §37 of Part V we have shown that a level corresponding to a bound state can be formed in a spherical potential well only under the condition that the depth of the potential well is larger than a certain critical value ((37.9) of Part V). In the present case a bound state can be formed when there is an arbitrarily small interaction (small F). Formally this is due to the fact that the summation in (65.2) is bounded by the condition $k > k_0$. Physically this is associated with the effect of background electrons.

Thus in a metal pairing of electrons may take place. The electron pairs have an integer spin and obey Bose statistics. An imperfect Bose gas possesses the property of superfluidity. Since the pairs are charged, the superfluid motion of electrons corresponds to the appearance of superconductivity. We note that the function (65.9) cannot be expanded in a series about the point $\beta = 0$ and, consequently, no calculations based on perturbation theory could lead to an explanation of the phenomenon of superconductivity.

The phenomenon of electron pairing is particularly clearly shown in the behaviour of doubly connected (hollow) superconductors in a magnetic field.

Let us consider a hollow superconducting cylinder placed in a uniform external magnetic field directed along its axis. We shall be interested in the wave function $\varphi_H(\mathbf{R})$ characterizing the motion of the centre of mass of an electron pair in the magnetic field. We need not take into account here the internal motion of the components of the pair.

Since the spin of the pair is equal to zero, and the charge and mass are twice as large as for one electron, the Schrödinger equation for the pair can be written in the form (see §102, Part V)

$$\frac{1}{4m} \left[\frac{\hbar}{i} \nabla - \frac{2e}{c} \mathbf{A}(\mathbf{r}) \right]^2 \varphi_H = E \varphi_H. \quad (65.11)$$

We note that when a magnetic field is applied, a superconducting current completely screening the external magnetic field arises at the surface of the

superconductor. Hence inside the superconductor

$$\mathbf{H} = \nabla \times \mathbf{A} = 0.$$

This means that the vector potential \mathbf{A} in a hollow cylinder can be written in the form

$$\mathbf{A} = \nabla \chi(\mathbf{r}), \quad (65.12)$$

where χ is the scalar potential of the magnetic field, which is an ambiguous function of the coordinates.

The magnetic flux through the cavity in the cylinder is equal to

$$\Phi = \int \mathbf{H} \cdot d\mathbf{S} = \int \nabla \times \mathbf{A} \cdot d\mathbf{S} = \oint \mathbf{A} \cdot d\mathbf{r} = \int \nabla \chi \cdot d\mathbf{r} = \Delta \chi, \quad (65.13)$$

where $\Delta \chi$ is the jump of the function Φ in going round a closed contour inside the superconducting cylinder.

Making use of expression (65.12), eq. (65.11) can be rewritten in the form

$$\frac{1}{4m} \left[\frac{\hbar}{i} \nabla - \frac{2e}{c} \nabla \chi(\mathbf{r}) \right]^2 \varphi_{\mathbf{H}} = E \varphi_{\mathbf{H}}. \quad (65.14)$$

It is easily seen that the wave function, $\varphi_{\mathbf{H}}$, of the pair in the magnetic field is expressed in terms of the wave function, φ_0 , of the pair in the absence of magnetic field by the relation

$$\varphi_{\mathbf{H}} = \varphi_0 e^{-(2ie/\hbar c)\chi(\mathbf{r})} = \frac{1}{V^{1/2}} e^{i\mathbf{K} \cdot \mathbf{r} - (2ie/\hbar c)\chi(\mathbf{r})}. \quad (65.15)$$

Of course, the wave function $\varphi_{\mathbf{H}}$ must be a single-valued function of coordinates, as must the wave function φ_0 .

Let us in imagination displace an electron pair inside the superconductor along an arbitrary closed curve. In such a displacement neither the function φ_0 nor the function $\varphi_{\mathbf{H}}$ can change. However, according to (65.15), the ambiguous function $\chi(\mathbf{r})$ changes by an amount $\Delta \chi$ in going round the closed contour. In order for the wave function $\varphi_{\mathbf{H}}$ not to change when χ increases by the amount $\Delta \chi$, it is necessary that the following equality be fulfilled:

$$\frac{2e}{\hbar c} \Delta \chi = \frac{2e\Phi}{\hbar c} = 2\pi n,$$

where $n = 0, \pm 1, \pm 2, \dots$. Thus the magnetic flux through the cavity of the cylinder can run over the discrete series of values

$$\Delta\Phi = \frac{\hbar c}{2e} 2\pi n, \quad (65.16)$$

which are multiples of the quantity $\hbar c/2e$ and depend on the charge ($2e$). The total magnetic flux through the cavity is obtained by multiplying $\Delta\Phi$ by the number of pairs in the volume of the superconducting cylinder.

The fact of quantization of the magnetic flux through the cavity of the cylinder, as well as its dependence on the pair charge ($2e$), have been experimentally confirmed.

It is useful to supplement the above qualitative consideration of the phenomenon of superconductivity by a more consequential discussion of the properties of a particular simplified model described by the Hamiltonian

$$H = \sum [\epsilon(p) - \mu] (\hat{a}_{\mathbf{p}\uparrow}^\dagger \hat{a}_{\mathbf{p}\uparrow} + \hat{a}_{\mathbf{p}\downarrow}^\dagger \hat{a}_{\mathbf{p}\downarrow}) + \Delta \sum (\hat{a}_{\mathbf{p}\uparrow} \hat{a}_{-\mathbf{p}\downarrow} + \hat{a}_{-\mathbf{p}\downarrow}^\dagger \hat{a}_{\mathbf{p}\uparrow}^\dagger), \quad (65.17)$$

where

$$\epsilon(p) = \frac{p^2}{2m} = vp$$

is the energy of an electron of momentum p , $\mu = p_0^2/2m$ is the chemical potential, and p_0 is the momentum at the Fermi surface. The first term of the Hamiltonian (65.17) represents the Hamiltonian of a system of non-interacting particles outside the Fermi surface. The second sum characterizes the interaction of electrons outside the Fermi surface with the condensate, i.e. with electrons in occupied states. This interaction leads to the creation or annihilation of pairs of electrons with oppositely directed momenta and spins.

For reasons which will become clear from what follows, the quantity Δ , called the energy gap, represents the work of removing a pair from the condensate to the Fermi surface.

In choosing the Hamiltonian (65.17) we have made use of the fact that the number of particles in the condensate is large. Hence, just as was done in §5, instead of the four-fermion interaction operator figuring in Fröhlich's Hamiltonian we have introduced simplified two-fermion operators: the operator $\hat{a}_{\mathbf{p}\uparrow} \hat{a}_{-\mathbf{p}\downarrow}$ describing the destruction of two electrons, one with momentum \mathbf{p} and spin up and the other with momentum $(-\mathbf{p})$ and spin down, and the operator $\hat{a}_{\mathbf{p}\uparrow}^\dagger \hat{a}_{-\mathbf{p}\downarrow}^\dagger$ describing the creation of such a pair.

We carry out a linear transformation on the Hamiltonian (65.17), passing from the operators \hat{a} and \hat{a}^\dagger to new operators $\hat{\alpha}$ and $\hat{\alpha}^\dagger$, analogous to that carried out in §5.

Namely, we set

$$\hat{a}_{\mathbf{p}\uparrow} = u_{\mathbf{p}} \hat{\alpha}_{\mathbf{p}\uparrow} + v_{\mathbf{p}} \hat{\alpha}_{-\mathbf{p}\downarrow}^\dagger, \quad (65.18)$$

$$\hat{a}_{\mathbf{p}\downarrow}^\dagger = u_{\mathbf{p}} \hat{\alpha}_{\mathbf{p}\downarrow}^\dagger - v_{\mathbf{p}} \hat{\alpha}_{-\mathbf{p}\uparrow}, \quad (65.19)$$

Here $u_{\mathbf{p}}$ and $v_{\mathbf{p}}$ are transformation coefficients which are assumed to be real c -numbers, so that

$$u_{\mathbf{p}} = u_{-\mathbf{p}}; \quad v_{\mathbf{p}} = v_{-\mathbf{p}}.$$

Furthermore, the coefficients $u_{\mathbf{p}}$ and $v_{\mathbf{p}}$ are assumed to satisfy the condition

$$u_{\mathbf{p}}^2 + v_{\mathbf{p}}^2 = 1. \quad (65.20)$$

The coefficients $u_{\mathbf{p}}$ and $v_{\mathbf{p}}$ are not completely defined by formulae (65.18) and (65.19), and we can subsequently impose upon them one more condition at will.

It is easily seen directly that transformations (65.18) and (65.19) are canonical, i.e. that the new operators $\hat{\alpha}$ and $\hat{\alpha}^\dagger$, as well as the old operators \hat{a} and \hat{a}^\dagger , satisfy the Fermi commutation relations. By substituting into the Hamiltonian (65.17) and transforming it to the new Fermi operators, we obtain, taking into account (65.20),

$$\begin{aligned} H = & 2 \sum [\epsilon(\mathbf{p}) - \mu] v_{\mathbf{p}}^2 - 2\Delta \sum u_{\mathbf{p}} v_{\mathbf{p}} + \\ & + \sum \{ [\epsilon(\mathbf{p}) - \mu] (u_{\mathbf{p}}^2 - v_{\mathbf{p}}^2) + 2\Delta u_{\mathbf{p}} v_{\mathbf{p}} \} (\hat{a}_{\mathbf{p}\uparrow}^\dagger \hat{a}_{-\mathbf{p}\downarrow} + \hat{a}_{\mathbf{p}\downarrow}^\dagger \hat{a}_{-\mathbf{p}\uparrow}) + \\ & + \sum \{ [\epsilon(\mathbf{p}) - \mu] 2u_{\mathbf{p}} v_{\mathbf{p}} - \Delta(u_{\mathbf{p}}^2 - v_{\mathbf{p}}^2) \} (\hat{a}_{\mathbf{p}\uparrow}^\dagger \hat{a}_{\mathbf{p}\uparrow}^\dagger + \hat{a}_{\mathbf{p}\downarrow} \hat{a}_{\mathbf{p}\downarrow}). \end{aligned}$$

We now require that the coefficients $u_{\mathbf{p}}$ and $v_{\mathbf{p}}$ satisfy the condition

$$2[\epsilon(\mathbf{p}) - \mu] u_{\mathbf{p}} v_{\mathbf{p}} - \Delta(u_{\mathbf{p}}^2 - v_{\mathbf{p}}^2) = 0.$$

Then the expression for H is essentially simplified and assumes the form

$$H = E_0 + \sum E(\mathbf{p}) (\hat{a}_{\mathbf{p}\uparrow}^\dagger \hat{a}_{-\mathbf{p}\downarrow} + \hat{a}_{-\mathbf{p}\downarrow}^\dagger \hat{a}_{\mathbf{p}\uparrow}), \quad (65.21)$$

where

$$E_0 = 2 \sum \{ [\epsilon(\mathbf{p}) - \mu] v_{\mathbf{p}}^2 - \Delta u_{\mathbf{p}} v_{\mathbf{p}} \}, \quad (65.22)$$

$$E(\mathbf{p}) = [\epsilon(\mathbf{p}) - \mu] (u_{\mathbf{p}}^2 - v_{\mathbf{p}}^2) + 2\Delta u_{\mathbf{p}} v_{\mathbf{p}}. \quad (65.23)$$

The Hamiltonian (65.21) has an obvious interpretation. It corresponds to a gas of elementary excitations. The elementary excitations possess energy $E(\mathbf{p})$ and are associated with the motion of two fermions with momenta (\mathbf{p}) and $(-\mathbf{p})$. The value of $E(\mathbf{p})$ can be expressed in terms of $\epsilon(\mathbf{p})$, μ and Δ , if eqs. (65.22) and (65.23) are solved for $u_{\mathbf{p}}$ and $v_{\mathbf{p}}$.

Obviously, from (65.22), (65.21) and (65.20) we have

$$u_{\mathbf{p}}^2 = \frac{1}{2} \left(1 + \frac{\epsilon(\mathbf{p}) - \mu}{\epsilon(\mathbf{p})} \right), \quad v_{\mathbf{p}}^2 = \frac{1}{2} \left(1 - \frac{\epsilon(\mathbf{p}) - \mu}{\epsilon(\mathbf{p})} \right),$$

Substituting the values of $u_{\mathbf{p}}$ and $v_{\mathbf{p}}$ into (65.23), we obtain for $E(\mathbf{p})$

$$E(\mathbf{p}) = (|\epsilon(\mathbf{p}) - \mu|^2 + \Delta^2)^{\frac{1}{2}}. \quad (65.24)$$

Thus the minimum energy of the elementary excitations (for $\epsilon(\mathbf{p}) = \mu$) turns out to be equal to Δ . A gap of width Δ separates the energy of an elementary excitation from the energy of particles at the Fermi surface. Near the Fermi surface, $E(\mathbf{p})$ can be written in the form

$$E(\mathbf{p}) = (v^2(\mathbf{p} - \mathbf{p}_F)^2 + \Delta^2)^{\frac{1}{2}}. \quad (65.25)$$

As we have seen in §5, the gas of elementary excitations with an energy spectrum given by formulae (65.21) and (65.25) possesses the property of superfluidity. In our case the motion of the elementary excitation is connected with the charge transfer. Therefore the gas of elementary excitations has the property of superconductivity.

As was shown in §5, the system of particles can move as a whole without any dissipation if the inequality

$$|\mathbf{v}^*| = \min \frac{\epsilon(\mathbf{p})}{|\mathbf{p}|} > 0 \quad (65.26)$$

is fulfilled. In our case this inequality is just fulfilled if the quantity Δ is posi-

tive: from formulae (65.24) and (65.25) we have

$$|\mathbf{v}^*| = \frac{\Delta}{p_F(1 + \Delta^2/v_0^2 p_F^2)^{1/2}} \simeq \frac{\Delta}{p_F}. \quad (65.27)$$

Therefore if the system of electrons moves as a whole with velocity less than Δ/p_F such motion will be nondissipative. In other words electrons will move through the crystalline body without any resistance.

It may be shown that the width of the gap Δ is dependent on the temperature and decreases as the latter increases. For some temperature T the width of the 'gap' Δ tends towards zero. Here the property of superconductivity disappears and a metal becomes a normal conductor.

It is necessary to emphasize that besides the effective attraction between electrons the usual Coulomb repulsion exists. If the latter dominates the system of electrons the electrons cannot be in a superconducting state.

Thus we see that all conclusions about the property of isolated Fermi systems remain the same. Such a system cannot have the property of superconductivity. A system of interacting electrons which possesses the collective states and collective motion only arises in a crystalline body where a mixture of electrons and photons exists: under certain conditions such a system can exist in a superconducting state.

It should also be noted that the interaction via pair exchange of phonons is apparently not the only type of interaction between electrons which is responsible for superconductivity. This is indicated by the fact that there are superconductors (ruthenium, osmium) which do not display the isotope effect. However, from the preceding calculations it is seen that the theory of superconductivity we have is not very sensitive to the detailed character of the interaction forces (attractive forces) between electrons which is responsible for the appearance of the property of superconductivity.

§66. Theory of the Fermi fluid

A few years ago Landau put forward a phenomenological theory of the Fermi fluid, in which the existence of a strong interaction between fermions is assumed from the very beginning. This theory subsequently obtained a statistical substantiation which, however, is too complex to be presented in this book†.

* A.A. Abrikosov, L.P. Gorkov and I.E. Dzyaloshinskii, *Methods of quantum field theory in statistical physics* (Prentice-Hall, Englewood Cliffs, N.Y., 1963). In presenting the theory of the Fermi fluid we follow this book.

The theory was applied to the description of the behaviour of the liquid helium isotope ^3He , whose nuclei have spin $\frac{1}{2}$ and obey Fermi statistics. If the assumption is made that the energy spectrum of a system of electrons in a crystal lattice differs relatively little from the spectrum of the electron fluid filling the corresponding volume, then it applies equally to electrons in metals.

The theory is based on the assumption that no matter how strong the interaction between the particles may be, it cannot violate the exclusion principle. Hence the occupation numbers of energy states in a liquid, as well as in a gas, can be equal only to either zero or one. This means that in a Fermi fluid at absolute zero all energy levels are occupied up to the Fermi boundary surface.

The energy distribution has the character of a step function

$$f = \begin{cases} 0 & \epsilon > \epsilon_F, \\ 1 & \epsilon < \epsilon_F. \end{cases}$$

The energy and momentum at the Fermi surface are connected with the number of particles by the relation (see §79 of Part III)

$$\epsilon_F = \frac{p_F^2}{2m} = \frac{(2\pi\hbar)^2}{2m} \left(\frac{3N}{8\pi V} \right)^{\frac{2}{3}}. \quad (66.1)$$

At a temperature different from zero but sufficiently low, collective excitations arise and the step-wise distribution is somewhat smeared. If the temperature is sufficiently low, then it can be assumed that the energy of these excitations is close to the limiting energy ϵ_F .

The appearance of excitations is always accompanied by the formation of vacancies (holes) in occupied states within the Fermi surface. The disappearance of excitations is associated with the occupation of vacant states, the annihilation of 'holes' and 'particles'. Hence excitations appear and disappear in pairs.

If the excitation energy is measured from the Fermi surface and the excitations are assigned a definite value of momentum, then they can be treated as a pair of quasiparticles; properly speaking a quasiparticle and a hole.

The excitation energy for small excitations can be written in the form of an expansion in terms of a small parameter $(p - p_F)$

$$\epsilon = \epsilon_F + \left(\frac{\partial \epsilon}{\partial p} \right)_{p_F} (p - p_F). \quad (66.2)$$

Analogously, the energy of a hole (measured, as always, downward from the Fermi surface) is

$$-\epsilon = \epsilon_F + \left(\frac{\partial \epsilon}{\partial p} \right)_{p_F} (p_F - p). \quad (66.3)$$

In accordance with the general propositions of the theory of quasiparticles, the two forms of particles are equivalent and their properties can be described by the quantity

$$\left(\frac{\partial \epsilon}{\partial p} \right)_{p=p_0} = \frac{p_0}{m^*}, \quad (66.4)$$

where m^* is the effective mass.

At low temperatures the width of the region of energies of excited particles is

$$\epsilon - \epsilon_F \sim kT. \quad (66.5)$$

For the approximation of free quasiparticles to make sense, it is necessary that the concept of momentum makes sense, i.e. that condition (47.19) be fulfilled.

Since excitations in the Fermi fluid may appear and disappear only in pairs, the number of encounters of a pair 'particle'-'hole' is proportional to N^2 , where N is the total number of excitations at a given temperature.

At low temperatures N increases proportionally to T . Correspondingly, the lifetime of the excitations is

$$\tau \sim \frac{1}{N^2} \sim \frac{1}{T^2}. \quad (66.6)$$

Hence $\Delta\epsilon$, the uncertainty in the energy of the quasiparticle, is, according to (66.6), equal to

$$\Delta\epsilon \sim \hbar/\tau \sim \hbar T^2. \quad (66.7)$$

Comparing (66.7) and (66.5) we see that for a sufficiently low temperature we always have

$$(\epsilon - \epsilon_F) \gg \Delta\epsilon, \quad (66.8)$$

and the concept of momentum for independent quasiparticles makes sense.

Thus quasiparticles in a Fermi fluid have momentum and effective mass. They appear or disappear only in pairs in rare collisions and, consequently, obey the exclusion principle.

The properties of elementary excitations enumerated allow one to treat them as a Fermi system described by the distribution function

$$f(\epsilon) = \frac{1}{\exp[(\epsilon - \epsilon_F)/kT] + 1}. \quad (66.9)$$

We shall normalize the distribution function of quasiparticles by the condition

$$\int f(\epsilon) \frac{d\Gamma}{(2\pi\hbar)^3} = \frac{N}{V}, \quad (66.10)$$

where N is the number of real particles of the fluid in volume V .

However, and here lies the essential difference between a Fermi liquid and a Fermi gas, the energy ϵ of a given quasiparticle depends on the density $n(\epsilon)$ and, consequently, also on the temperature.

In the approximation of the self-consistent field each quasiparticle moves in the field produced by all the other quasiparticles. Hence, if the distribution function of the excitations changes by amount δf , then the energy of the quasiparticle will change correspondingly. This change can be written as

$$\delta\epsilon = \int \hat{F}(\mathbf{p}, \mathbf{p}') \delta f(\mathbf{p}') \frac{d\mathbf{p}'}{(2\pi\hbar)^3}, \quad (66.11)$$

where the operator $\hat{F}(\mathbf{p}, \mathbf{p}')$ describes the change we are interested in. (For simplification of the formulae we do not write here the dependence of the operator \hat{F} on spins.) This operator, introduced formally, plays the principal role in the theory of the Fermi fluid. Formula (66.11) shows that the energy ϵ is a functional of the distribution f .

The properties of the Fermi distribution determine the temperature dependence of the thermodynamic characteristics of the ideal gas. Namely, using the formulae of §80 of Part III, we have

$$E = \int \frac{\epsilon}{e^{(\epsilon - \mu)/kT} + 1} \frac{d\Gamma}{(2\pi\hbar)^3} \simeq \frac{3}{5} N \epsilon_F \left[1 + \frac{5\pi^2}{12} \left(\frac{kT}{\epsilon_F} \right)^2 \right] \quad (66.12)$$

and

$$C_V = kN \frac{\pi^2 kT}{\epsilon_F}. \quad (66.13)$$

The linear temperature dependence of the heat capacity is associated solely with the step-like character of the Fermi distribution.

The kinetic properties of a Fermi fluid turn out to be more remarkable. According to (66.6) the mean free path of quasiparticles is

$$\lambda \sim v_F \tau \sim \frac{v_F}{T^2} \quad (66.14)$$

and increases rapidly with decreasing temperature. This fact determines the temperature dependence of the kinetic coefficients which increase with decreasing temperature. For example, the viscosity in order of magnitude is

$$\eta \sim (m^* v_F) n \lambda \sim \frac{1}{T^2} \quad (66.15)$$

and, correspondingly, the thermal conductivity is

$$\kappa \sim (C_V n) v_F \lambda \sim \frac{1}{T}. \quad (66.16)$$

The large mean free path of quasiparticles makes the propagation of ordinary sound in a Fermi fluid impossible. As we have seen in §26, for wavelengths exceeding the mean free path length, very strong damping of sound waves occurs in the medium.

However, it turns out that periodic perturbations of high frequency $\omega > 1/\tau$, having a character essentially different from that of ordinary sound waves, may propagate in a Fermi fluid.

Let us write the kinetic equation for the non-equilibrium distribution function

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{F} \cdot \frac{\partial f}{\partial \mathbf{p}} = I \quad (66.17)$$

and substitute in it

$$\mathbf{F} = - \frac{\partial \epsilon}{\partial \mathbf{r}}. \quad (66.18)$$

Then the kinetic equation assumes the form

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} - \frac{\partial \epsilon}{\partial \mathbf{r}} \cdot \frac{\partial f}{\partial \mathbf{p}} = I. \quad (66.19)$$

The collision integral is $I \sim 1/\tau$ and for sufficiently low temperatures is small. In other words, at sufficiently low temperatures collisions between quasiparticles become so rare that they do not affect the distribution function much.

Under these conditions the transmission of perturbations in an ideal gas is stopped. As can be seen from (66.12), the interaction existing in the Fermi fluid leads to a change in the energy of quasiparticles as the distribution function changes. This is a specific form of the long-range action between particles in a Fermi fluid. Owing to the long-range forces, a perturbation of the distribution function arising at a particular instant of time propagates throughout the fluid.

In the kinetic equation we set

$$f = f_0(\epsilon) + f_1(\mathbf{p}), \quad (66.20)$$

where $f_0(\epsilon)$ is the equilibrium distribution, and f_1 is the perturbation.

Substituting (66.20) into the kinetic equation, we find

$$\frac{\partial f_1}{\partial t} + \mathbf{v} \cdot \frac{\partial f_1}{\partial \mathbf{r}} - \frac{\partial f_0}{\partial \epsilon} \frac{\partial \epsilon}{\partial \mathbf{p}} \cdot \frac{\partial \epsilon}{\partial \mathbf{r}} = 0. \quad (66.21)$$

We note that in an equilibrium fluid $\partial \epsilon / \partial \mathbf{r} = 0$. Hence in the last term we have retained only small quantities of the first order.

Since the equilibrium distribution f_0 is of the form of a step function, we have (see § 80 of Part III)

$$\frac{\partial f_0}{\partial \epsilon} = -\delta(\epsilon - \epsilon_F).$$

The perturbation f_1 also must be proportional to this delta function. We shall seek it in the form

$$f_1 \sim \alpha \delta(\epsilon - \epsilon_F) e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}. \quad (66.22)$$

Choosing the vector \mathbf{k} to be the polar axis and assuming that α depends only on the angle θ , we obtain

$$i(v_F k \cos \theta - \omega) \alpha + \mathbf{v} \cdot \frac{\partial \epsilon}{\partial \mathbf{r}} = 0. \quad (66.23)$$

Making use of (66.11), one can write for the change of energy $\partial\epsilon/\partial\mathbf{r}$

$$\frac{\partial\epsilon}{\partial\mathbf{r}} = \int F(\mathbf{p}, \mathbf{p}') \frac{\partial f_1(\mathbf{p}')}{\partial\mathbf{r}} \frac{d\mathbf{p}'}{(2\pi\hbar)^3} = ik \int F(\mathbf{p}, \mathbf{p}') f_1(\mathbf{p}') \int \frac{d\mathbf{p}'}{(2\pi\hbar)^3}. \quad (66.24)$$

In substituting (66.24) into eq. (66.23) one has to take the values of momentum at the Fermi surface, so that the integration over $d\mathbf{p}$ is transformed to an integration over angles. We thus obtain

$$(kv_F \cos\theta - \omega) \alpha(\theta) + \frac{kv_F \cos\theta p_F^3}{3(2\pi\hbar)^3} \int F \alpha(\theta') d\Omega = 0. \quad (66.25)$$

The dependence of F on the angle is unknown. If it is assumed that F does not depend on angles at all and that it is a constant, then

$$\frac{p_F^3}{3(2\pi\hbar)^3} F \int \alpha(\theta') d\Omega = \text{const} = A$$

and

$$\left(\frac{\omega}{kv_F} - \cos\theta \right) \alpha(\theta) = A \cos\theta$$

or

$$\alpha(\theta) = A \cos\theta / \left(\frac{\omega}{kv} - \cos\theta \right). \quad (66.26)$$

Substituting this value of α into (66.25), we obtain the equation for defining $\omega/kv = u/v$, where u is the velocity of propagation of perturbations

$$uv_F^{-1} \ln \left(\frac{uv_F^{-1} + 1}{uv_F^{-1} - 1} \right) = 1 + \frac{1}{A} = 1 + \frac{3(2\pi\hbar)^3}{2\pi F p_F^3}. \quad (66.27)$$

From this expression it is seen directly that the velocity of propagation of perturbations can be real only if it exceeds the velocity of particles at the Fermi surface $v_F = p_F/m^*$. Then, as is seen from (66.22), the perturbed distribution function turns out to be distended in the direction of propagation of the perturbation.

The result obtained, as may be shown by a more detailed treatment, is of a general character and is not associated with the assumption that $F = \text{const.}$ Taking into account the angular dependence of F leads to an even more asymmetric distribution of the perturbed function.

The propagation of the perturbations considered is a specific effect, associated with the interaction of quasiparticles. This effect is called 'zero sound'. As is seen from the foregoing, zero sound represents an essentially non-equilibrium process. A more complete calculation shows that the propagation of zero sound is accompanied by a rapid damping over a length $\sim u\tau$.

The existence of 'zero sound' has recently been completely confirmed by experiment. The propagation and damping of zero sound excited by oscillations of the walls of a container were observed in liquid helium ^3He filling the container. The experiment confirmed all the basic conclusions of the theory.

The properties of the electron Fermi fluid turn out to be in many respects similar to those of a Fermi fluid of neutral particles. This applies to the general character of the onset of excitations, to the form of the energy spectrum, etc. At the same time the presence of the Coulomb interaction and the interaction with lattice phonons leads to the appearance of essential, although quantitative differences. The theory of the electron Fermi fluid is too complex and insufficiently developed to be discussed within the framework of this book.

§67. Electrons in dielectric crystals

As we have already stressed, the difference between metals, semiconductors and dielectrics is associated mainly with the different character of the electron energy level spectrum.

In dielectric crystals the empty band is separated from the filled band by a broad (of the order of one or several electron volts) band of forbidden energies.

Suppose that the absorption of light takes place with the excitation of one of the lattice atoms in a dielectric. Since the crystal possesses translational symmetry and the wave function of the excited state ψ_n must satisfy the condition of translational symmetry

$$T\psi_n = \alpha_n \psi_n, \quad (67.1)$$

the excited state cannot be localized at a definite atom. On the contrary, it

must move across the crystal and represents an excited state of the crystal as a whole.

The electron wave function describing the crystal in an excited state can be written in the form (see eq. (49.11))

$$\Psi = \sum e^{ik \cdot r} \varphi_j, \quad (67.2)$$

where φ_j is the symmetrized wave function corresponding to the excitation of the j th atom in the lattice. The wave function (67.2) satisfies both the Schrödinger equation (for proper choice of $E(\mathbf{k})$) and the requirement of translational symmetry (67.1).

We see that a collective excitation, an excitation transmitted in a lattice from one atom to another, can be considered as a quasiparticle with quasi-momentum $\mathbf{p} = \hbar \mathbf{k}$. This quasiparticle is called an exciton. All that was said in §47 on the properties of quasiparticles applies to the exciton. Since excitons appear and disappear one at a time, they must be considered as particles obeying the Bose–Einstein statistics.

In contrast to phonons, which represent collective excitations of the vibrational states of the nuclei of a lattice, excitons represent an excited electron state. However, the displacement of an exciton in a crystal is not associated with the motion of an electric charge. Herein lies the difference between an exciton and a charged quasiparticle, an electron or hole in a metal or semiconductor. The energy spectrum of an exciton $\epsilon(\mathbf{k})$ can be found in two limiting cases.

The first of these is the case of the so-called exciton of small radius or the Frenkel exciton. To such an exciton there corresponds an excited state of the crystal which is like an ordinary excitation in an isolated atom. This means that the excited electron is in the main localized near a certain atom of the lattice. Owing to the interaction between atoms, the excitation is transmitted to neighbouring atoms and thus migrates through the crystal. The Frenkel exciton can be pictured simply as an electron–ion residue pair whose radius is small in comparison with the lattice constant a .

The wave function φ_j of an exciton of small radius is defined by the Schrödinger equation with the Coulomb interaction potential

$$u = -e^2/r \quad (67.3)$$

between the electron and ion residue.

The energy spectrum of an exciton of small radius can be found in the approximation of interaction with nearest neighbours.

The corresponding calculations do not differ from those of §49 and lead to an expression for $\epsilon(\mathbf{k})$ which is in principle the same as (49.13) (we recall that the specific form of formula (49.13) refers to an electron in the s-state; this assumption is often unjustified for excited states). The spectrum of excitons has a band character. At the edges of the band the energy is connected with the momentum by the relation

$$\epsilon_n(\mathbf{k}) \sim \frac{\hbar^2 k^2}{2m^*}, \quad (67.4)$$

where m^* is expressed in terms of the exchange integral of interaction between the atoms, and n is the number of the band. In order of magnitude the effective mass of an exciton m^* is close to the analogous quantity for strongly bound electrons.

The other limiting case is that of an exciton of large radius or the Wannier exciton. The production of an exciton of large radius corresponds to the ionization of an isolated atom. This means that the excitation energy is sufficiently large to bring the electron from the filled into the vacant band of the crystal. Such a transition is accompanied by the simultaneous appearance of a hole in the filled band. The electron and hole interact with each other according to the Coulomb law in a medium

$$U = - \frac{e^2}{\epsilon_\infty |\mathbf{r}_e - \mathbf{r}_p|}, \quad (67.5)$$

where ϵ_∞ is the dielectric constant of the medium. Somewhat later we shall define what is meant exactly by ϵ_∞ .

Formula (67.5) with the macroscopic constant ϵ_∞ means that the interaction of the electron and hole is described by the averaged field of a rather large number of atoms and, consequently, that the distance between the components of the pair is rather large in comparison with the lattice constant.

It is the interacting electron-hole pair in a crystal that is called an exciton of large radius. The Schrödinger equation for an exciton, i.e. for a two-particle system, can be written in the form of (14.5) of Part V.

Introducing the radius vector of the center of mass of relative motion, its solution can be written in the form (see (14.14) of Part V)

$$\Psi = \psi \varphi$$

where φ satisfies the Schrödinger equation for relative motion

$$\left(-\frac{\hbar^2}{2\mu}\nabla_r^2 - \frac{e^2}{\epsilon_{\infty}r}\right)\varphi = \epsilon\varphi, \quad (67.6)$$

and ψ satisfies the equation of motion of the centre of mass in the crystal

$$\left(-\frac{\hbar^2}{2M}\nabla_R^2 + U\right)\psi = (E + \epsilon)\psi. \quad (67.7)$$

Eq. (67.6) is the same as the Schrödinger equation for the hydrogen atom, in which $Z = 1/\epsilon_{\infty}$. Its solution has the form of the hydrogen functions φ with energies ϵ_n , which in the region of the discrete part of the spectrum are given by formula (38.17) of Part V

$$\epsilon_n = -\frac{\mu e^2}{2\hbar^2 \epsilon_{\infty}^2 n^2}. \quad (67.8)$$

The solution of the Schrödinger equation for the motion of the centre of mass is the function (67.2).

The total energy of an exciton forms a band. At the edge of the band it can be written in the form

$$E_{n,k} = \frac{\hbar^2 k^2}{2\mu} - \frac{\mu e^2}{2\hbar^2 \epsilon_{\infty}^2 n^2}. \quad (67.9)$$

Thus an exciton of large radius moves in the crystal as a whole and at the same time has an internal degree of freedom.

A remarkable feature of an exciton of large radius is just the presence in its energy spectrum of a hydrogen-like term corresponding to the internal degree of freedom. In the spectrum of absorption of light in a dielectric, discrete levels corresponding to the hydrogen spectrum must appear. This effect has been observed in a number of crystals, and can be considered as the most direct proof of the existence of excitons of large radius. The size of a large exciton, r_W , is defined by the value of the Bohr radius for $Z = 1/\epsilon_{\infty}$, i.e. is given by the formula

$$r_W = \frac{\hbar^2 \epsilon_{\infty}}{\mu e^2}.$$

For the concept of the exciton of large radius to make sense, it is necessary that the following inequality be fulfilled

$$\frac{\hbar^2 \epsilon_\infty}{\mu e^2} > a. \quad (67.10)$$

We note that besides discrete states, eq. (67.6) has continuous spectrum solutions. These states correspond to the disintegration of the exciton into an electron and a hole moving independently of each other in the crystal. In contrast to the motion of an exciton, in this case not only energy transport but also charge transport (photoconductivity of the dielectric, if the excitation is produced by irradiation) arises.

From the above it is clear that the difference between excitons of large and small radii is of a somewhat relative character. Intermediate cases are possible, where the size of the exciton is comparable with the lattice constant.

We can now discuss what value of dielectric constant of the medium should be inserted in formula (67.5). The electron and hole rotate about their common centre of mass with a velocity very large compared with the velocity of motion of heavy nuclei. At the same time their velocity is comparable with the velocities of valence electrons. Hence valence electrons have time to become polarized, and their polarization follows the motion of the component of the pair, whereas the nuclei remain at rest. This situation corresponds to the action of a high-frequency electromagnetic field on the crystal. Hence ϵ_∞ should be understood to be n^2 , the dielectric constant at high frequencies.

In conclusion we note that the concept of an exciton has no meaning for metals. In metals the lifetime, Δt , of excitons turns out to be so small that the concept of a quasiparticle makes no sense:

$$\Delta \epsilon \sim \frac{\hbar}{\Delta t} \sim \epsilon_n.$$

Superconductors, having a gap in the energy spectrum, could be an exception to this.

§68. The energy spectrum and the distribution function of electrons in semiconductors

Semiconductors, which in their electronic properties occupy an intermediate place between metals and insulators, have assumed very great importance in modern physics and technology.

The electronic properties of semiconductors, as well as analogous properties of metals and insulators, are defined in the first place by the character of their energy spectrum. Let us consider a body whose lower energy band is completely filled, at absolute zero, whereas the upper one is completely empty. We assume, however, that the spacing $\Delta\epsilon$ between the filled and empty bands is very small. In such a body there is a certain probability of exciting an electron and of its transition into the empty band. At a temperature $T > \Delta\epsilon/k$ the thermal energy is comparable with the width of the gap $\Delta\epsilon$, and the transition probability becomes of the order of unity. An electron which gets into the empty band can move under the action of an external field and can take part in conduction.

In addition to thermal excitation, other mechanisms giving rise to the transition of electrons from the filled into the empty energy band may exist. As an example photoexcitation can be mentioned, in which the necessary energy is imparted to electrons by irradiating the sample.

The transfer of electrons from the filled band is accompanied by the appearance of vacancies in it. The presence of vacancies allows the possibility of electrons of the filled band changing energy states. A vacancy in the filled band is called a hole, and the above mentioned process of filling vacancies is said to be the motion of holes. In an external field the electrons transfer to the conduction band and the holes in the filled band move.

Bodies with such an electron spectrum are called intrinsic semiconductors.

Semiconductors containing impurities of other substances and thus called extrinsic semiconductors play a particularly important role in modern physics.

Let the atoms of the impurities be introduced into the crystal lattice of a semiconductor. The number of these atoms is considered sufficiently small to make it possible to disregard the interaction between them.

Suppose the energy level arising from the impurity lies very close to the lower edge of the empty band. Then at $T \neq 0$ the thermal ionization of the atoms of the impurity will occur with high probability. We stress that the thermal ionization of the atoms of the impurity in semiconductors occurs at substantially lower temperatures than thermal ionization in gases. The latter usually occurs only at temperatures of the order of several thousand degrees, when the energy of thermal motion becomes comparable with the binding energy of electrons in atoms (the ionization energy). In semiconductors the situation is fundamentally different, since in a continuous medium with dielectric constant ϵ the ionization energy decreases by a factor of ϵ^2 , i.e. 150–250 times. Thus, for example, the ionization energy of hydrogen atoms in germanium falls to about $0.016 \text{ eV} \approx 150 \text{ K}$.

The impurity atoms which provide electrons in the empty band are called

donors, and the semiconductor described is said to be an n-type extrinsic semiconductor. The index n emphasizes the fact that there are only negative free charges in the semiconductor.

If the impurity atoms have an affinity for electrons and easily form negative ions, the energy level arising from the impurity lies very close to the upper edge of the valence band. Then electrons from the filled band pass to the levels of the impurity atoms. Such transitions are carried out at $T \neq 0$ with considerable probability, if the spacing between the energy levels of the impurity and the upper edge of the filled band is sufficiently small.

The vacancies in the filled band ensure the possibility of motion of the holes in it and thus of charge transport in an external electric field. The impurity atoms which can capture electrons from the filled band are called acceptors. Semiconductors containing an acceptor impurity are called p-type extrinsic semiconductors. The index p emphasizes that the conductivity of the semiconductor is associated with the motion of positive holes in it (hole-type conductivity).

In real semiconductors the intrinsic conductivity is very often comparable with the extrinsic conductivity (n- or p-type); the semiconductor contains both acceptor and donor impurities. Finally, a very important role is played by spatially nonuniform conductors in which one part contains mainly an acceptor impurity, whereas the other part contains mainly a donor impurity. However, we shall confine ourselves to the consideration of the three limiting cases mentioned.

In the case of a semiconductor with intrinsic conductivity, one can write the current density in the form

$$\mathbf{j} = -e \sum_i f_i \mathbf{v}_i - e \sum_k f'_k \mathbf{v}_k, \quad (68.1)$$

where the summation over i refers to the quantum states of the empty band, and the summation over k to the states of the filled band.

We can write (68.1) in the form

$$\begin{aligned} \mathbf{j} &= -e \sum_i f_i \mathbf{v}_i - e \sum_k \mathbf{v}_k + e \sum_k (1 - f'_k) \mathbf{v}_k = \\ &= -e \sum_i f_i \mathbf{v}_i + e \sum_k (1 - f'_k) \mathbf{v}_k = \\ &= -e \sum_i f_n^{(i)} \mathbf{v}_i + e \sum_k f_p^k \mathbf{v}_k. \end{aligned} \quad (68.2)$$

The sum $e \sum_k v_k$ is equal to zero, since in the filled band the number of particles moving with the field is equal to that moving against the field. Formula (68.2) shows that the total current density is composed of the current of carriers in the empty band and the current of carriers in the filled band. The first are electrons with charge $(-e)$, the second are holes with effective charge $(+e)$.

We have denoted the distribution function of the electrons of the empty band by f_n , and the distribution function of holes in the filled band by f_p .

To find the probability of finding an electron in the state with energy ϵ one can write the Fermi distribution

$$f_n = \frac{1}{\exp [(\epsilon - \mu)/kT] + 1}. \quad (68.3)$$

Analogously, for the probability that an electron is absent from the state ϵ , i.e. that there is a hole, we have

$$f_p = 1 - f_n = \frac{1}{\exp [(\mu - \epsilon)/kT] + 1}. \quad (68.4)$$

Introducing the new variable $\epsilon' = -\epsilon - \Delta E$, where ΔE is the width of the forbidden zone, and measuring the energy from the lower edge of the conduction band, we rewrite (68.4) in the form

$$f_p = \frac{1}{\exp [(\epsilon' + \mu + \Delta E)/kT] + 1}.$$

The value of the chemical potential μ (often called the Fermi level) is defined by the requirement of the principle of detailed balance

$$\int f_n dp = \int f_p dp' \quad (68.5)$$

so that the semiconductor as a whole remains electrically neutral.

In most intrinsic semiconductors the number of electrons in the empty band is so small that the electron gas is non-degenerate. Hence one can write

$$f_n = e^{(\mu - \epsilon)/kT}. \quad (68.6)$$

Correspondingly, for the distribution function of holes we have

$$f_p = e^{-(\mu + \Delta E + \epsilon')/kT}. \quad (68.7)$$

Substituting (68.7) and (68.6) into (68.5), we find

$$\frac{2e^{\mu/kT}}{h^3} \int e^{-\epsilon/kT} 4\pi p^2 dp = \frac{2}{h^3} e^{-(\mu+\Delta E)/kT} \int e^{-\epsilon'/kT} 4\pi p'^2 dp'$$

or

$$\mu = -\frac{1}{2}\Delta E - \frac{1}{2}kT \ln(m_n/m_p)^{\frac{3}{2}}. \quad (68.8)$$

Correspondingly, the distribution functions of electrons and holes take the forms

$$f_n = \left(\frac{m_p}{m_n}\right)^{\frac{3}{4}} e^{-\Delta E/2kT} e^{-\epsilon/kT}, \quad (68.9)$$

$$f_p = \left(\frac{m_n}{m_p}\right)^{\frac{3}{4}} e^{-\Delta E/2kT} e^{-\epsilon'/kT}. \quad (68.9')$$

Here we have assumed that the masses of the electrons in the conduction band and the masses of holes have different values m_n and m_p .

Let us now turn to the case of an extrinsic semiconductor. For definiteness we shall assume that the semiconductor contains impurity atoms of only the donor type.

Let N_D be the total number of such atoms per unit volume of the semiconductor, and $(-E_D)$ the energy level measured from the lower edge of the conduction band. Then the condition for electric neutrality can be written in the form

$$N_D = \int_{-\infty}^{\infty} f_n \frac{dp}{d\epsilon} d\epsilon, \quad (68.10)$$

where the integration is carried out over the conduction band ($\epsilon > 0$) as well as over the impurity atoms.

For the latter one can obviously write

$$\frac{dp}{d\epsilon} \approx N_D \delta(\epsilon + E_D),$$

where $\delta(\epsilon + E_D)$ is the delta-function of the energy. Then

$$\begin{aligned}
 N_D &= N_D \int_{-\infty}^0 [e^{(\epsilon-\mu)/kT} + 1]^{-1} \delta(\epsilon + E_D) d\epsilon + \\
 &+ \int_0^{\infty} [e^{(\epsilon-\mu)/kT} + 1]^{-1} \frac{dp}{d\epsilon} d\epsilon = \\
 &= N_D [\exp \{-(E_D + \mu)/kT\} + 1]^{-1} + \frac{8\pi}{h^3} \int_0^{\infty} (e^{(\epsilon-\mu)/kT} + 1)^{-1} p^2 dp .
 \end{aligned} \tag{68.11}$$

If the number of electrons in the conduction band is small enough so that the electron gas may be considered non-degenerate, then

$$\begin{aligned}
 N_D &= N_D [\exp \{-(E_D + \mu)/kT\} + 1]^{-1} + \frac{8\pi e^{\mu/kT}}{h^3} \int_0^{\infty} e^{-\epsilon/kT} p^2 dp = \\
 &= N_D [\exp \{-(E_D + \mu)/kT\} + 1]^{-1} + 2 \left(\frac{2\pi mkT}{h^2} \right)^{\frac{3}{2}} e^{\mu/kT} .
 \end{aligned} \tag{68.12}$$

The condition (68.12) is quadratic with respect to $e^{\mu/kT}$ and makes it possible to express the chemical potential μ easily in terms of the characteristic quantities E_D and N_D . Solving eq. (68.12) gives

$$\mu = -E_D + kT \ln \frac{1}{2} \{ [1 + 2N_D h^3 \exp(E_D/kT)/(2\pi mkT)^{\frac{3}{2}}]^{\frac{1}{2}} - 1 \} .$$

For

$$\frac{N_D}{(2\pi mkT/h^2)^{\frac{3}{2}}} \exp(E_D/kT) \gg 1$$

we have

$$\mu = -\frac{1}{2}E_D + \frac{1}{2}kT \ln \frac{N_D}{2(2\pi mkT/h^2)^{\frac{3}{2}}} . \tag{68.13}$$

Hence we obtain for the distribution function of electrons in the conduction band

$$f_n = \frac{N_D}{2(2\pi mkT/\hbar^2)^{\frac{3}{2}}} \exp\left(-\frac{E_D + 2\epsilon}{2kT}\right). \quad (68.14)$$

The number of conduction electrons increases with temperature according to the law $\exp(-E_D/2kT)$.

When the inverse inequality

$$\frac{N_D}{(2\pi mkT/\hbar^2)^{\frac{3}{2}}} \exp(E_D/kT) \ll 1$$

is fulfilled, it can easily be shown that

$$\mu \approx kT \ln \frac{N_D}{2(2\pi mkT/\hbar^2)^{\frac{3}{2}}}. \quad (68.15)$$

It can easily be seen that this case corresponds to the position of the Fermi level (μ) being below the impurity level. Then all the atoms of the impurity must be fully ionized. The distribution function of electrons in the conduction band takes the form

$$f_n = \frac{N_D}{2(2\pi mkT/\hbar^2)^{\frac{3}{2}}} e^{-\epsilon/kT}. \quad (68.16)$$

The total number of electrons in the conduction band is then equal to N_D , so that there is saturation. A further increase in the number of free electrons with increasing temperature is possible only by means of the transfer of electrons from the filled band.

The distribution functions (68.15) and (68.16) contain the unknown quantities N_D , E_D and m , which are found from experimental data.

In the case of impurities of the acceptor type, all the calculations are carried out in exactly the same way. The resulting formulae are obtained by the replacement $N_D \rightarrow N_A$, $E_D \rightarrow E_A$, where the index A refers to the acceptor.

If the semiconductor contains acceptor and donor impurities at the same time, and if the width of the forbidden zone $\Delta\epsilon$ is comparable with E_D or E_A , then cumbersome expressions are obtained for the distribution functions. However, in principle they do not differ from the limiting expressions considered.

§69. The electrical conductivity and the Hall effect in semiconductors

To calculate the electrical conductivity and the Hall constant in semiconductors use can be made of the corresponding formulae of §58 and §59, which are of a general character. The specific characteristic of semiconductors is the possibility of the existence of current carriers of two signs. Furthermore, in most cases the electron gas and the hole gas in a semiconductor can both be considered non-degenerate. Then (58.3) gives

$$j = -CE \frac{n_{\text{eff}}}{2(2\pi mkT/\hbar^2)^{\frac{1}{2}}} \lambda_{\text{tr}} \epsilon \frac{\partial}{\partial \epsilon} e^{-\epsilon/kT} d\epsilon.$$

If the mean free path λ_{tr} is assumed to be independent of energy and equal to λ_0 , then

$$j = Ce \frac{n_{\text{eff}}}{2(2\pi mkT/\hbar^2)^{\frac{1}{2}}} \lambda_0 kT = \frac{4}{3} \frac{e^2 \lambda_0 n_{\text{eff}} E}{(2\pi mkT)^{\frac{1}{2}}}. \quad (69.1)$$

Correspondingly, we find for the electrical conductivity of a semiconductor

$$\sigma = \frac{4}{3} \frac{e^2 \lambda_0 n_{\text{eff}}}{(2\pi mkT)^{\frac{1}{2}}}. \quad (69.2)$$

Since n_{eff} depends on the temperature basically according to an exponential law, the electrical conductivity in semiconductors will be

$$\sigma \sim \exp(-\text{const}/T). \quad (69.3)$$

In front of the exponent there are factors containing different powers of the temperature. These powers are different for intrinsic and extrinsic semiconductors. However, they only affect the general temperature dependence of σ slightly.

We can also write the conductivity in the form

$$\sigma = \frac{\alpha e^2 \lambda_0 n_{\text{eff}}}{mv} = \frac{\alpha e^2 \tau}{m} n_{\text{eff}}, \quad (69.4)$$

where v is the mean velocity of thermal motion, $\tau = \lambda_0/v$ is the relaxation time, and α is a numerical coefficient.

In the case of two kinds of carriers, electrons and holes, the expression for

electrical conductivity can be written

$$\sigma = \frac{\alpha_1 e^2 \lambda_{el}^0 n_{eff}^{(el)}}{m_{el} v} + \frac{\alpha_2 e^2 \lambda_{hole}^0 n_{eff}^{(hole)}}{m_{hole} v} = \frac{\alpha_1 e^2 n_{eff}^{(el)}}{m_{el}} \tau_{el} + \frac{\alpha_2 e^2 n_{eff}^{(hole)}}{m_{hole}} \tau_{hole} . \quad (69.5)$$

Measurements of the Hall effect play an important role in determining the sign of current carriers in semiconductors. In calculating the Hall constant use can again be made of the general formulae of § 59. From (59.14) we find

$$R = \frac{L_2}{L_1} \frac{1}{H \sigma(H)} \quad (69.6)$$

where

$$\frac{L_2}{L_1} = \int_0^\infty \frac{\lambda \epsilon (\lambda \omega_H / v)}{1 + (\lambda \omega_H / v)^2} \frac{\partial f_0}{\partial \epsilon} d\epsilon \bigg/ \int_0^\infty \frac{\lambda \epsilon}{1 + (\lambda \omega_H / v)^2} \frac{\partial f_0}{\partial \epsilon} d\epsilon .$$

For real fields in semiconductors this relation can always be considered to be equal to

$$\begin{aligned} \frac{L_2}{L_1} &= \left[\int_0^\infty \lambda \epsilon \frac{\lambda \omega_H}{v} \frac{\partial f_0}{\partial \epsilon} d\epsilon \right] \bigg/ \left[\int_0^\infty \lambda \epsilon \frac{\partial f_0}{\partial \epsilon} d\epsilon \right] = \\ &= \frac{\lambda_0 e H}{mc} \left[\int_0^\infty \frac{\epsilon}{v} \frac{\partial f_0}{\partial \epsilon} d\epsilon \right] \bigg/ \left[\int_0^\infty \epsilon \frac{\partial f_0}{\partial \epsilon} d\epsilon \right] = \\ &= \frac{e H \lambda_0}{4mc} \left(\frac{2\pi m}{kT} \right)^{\frac{1}{2}} . \end{aligned} \quad (69.7)$$

In the first approximation, which is well fulfilled for semiconductors,

$$\sigma(H) = \sigma(0) = \sigma ,$$

where $\sigma(0) = \sigma$ is the electrical conductivity in the absence of a field as given by formula (69.2).

From (69.6), (69.7) and (69.2) we find the following value for the Hall

constant:

$$R = -\frac{3\pi}{8} \frac{1}{n_{\text{eff}} e c}. \quad (69.8)$$

This is the value of R for a semiconductor with one type of current carrier. If both electrons and holes take part in the conductivity, then, obviously, one can write for R

$$R = \frac{3\pi}{8c|e|} \left(\frac{1}{n_{\text{eff}}^{\text{el}}} - \frac{1}{n_{\text{eff}}^{\text{hole}}} \right). \quad (69.9)$$

The effective numbers of electrons and holes depend on the temperature, so that the Hall constant is not only a function of the temperature, but under certain conditions may change sign.

The relations obtained are valid only for semiconductors in which the number of carriers per cm^3 is small and so can be considered as a classical ideal gas. For some semiconductors, in particular for germanium, this assumption is not fulfilled. In this case the carriers must be considered as a weakly degenerate Fermi gas. In other words, the deviation of the distribution function from Maxwellian must be taken into account. We shall not dwell on a consideration of such semiconductors. Neither can we elucidate a number of effects, very important in practice, associated with the behaviour of semiconductors which are non-uniform in their properties. The theory of these effects is given in the specialist literature*.

* See R.D.Middlebrook, *An introduction to junction transistor theory* (Wiley, New York, 1957); W.Shockley, *Electrons and holes in semiconductors, with application to transistor electronics* (Van Nostrand, New York, 1950); A.F.Ioffe, *Physik der Halbleiter* (translation from Russian) (Akademie-Verlag, Berlin, 1958).

Interaction of Radiation with a Free-Electron Gas

§ 70. Low-density plasma in a low-frequency radiation field

Recent years have seen the discovery of a number of important astronomical objects that are the sources of extremely intense radiation. It turns out that the characteristic radiation spectrum of these objects is essentially different from the equilibrium (Planck) distribution, and has its intensity maximum in the region of radio or infrared waves. These objects, moreover, are surrounded by clouds (atmospheres) of a very low density ionised plasma. Apart from these powerful extra-terrestrial sources, radiating systems of a similar type have been developed under terrestrial conditions. These include lasers and modern UHF transmitters, where high-intensity radiation is created in the optical or radio wavebands.

Therefore the behaviour of a low-density plasma in a field of high-density radiation has lately caused a great deal of interest. It emerged that not even such a simple system as that of free electrons in a radiation field was paid sufficiently close attention. On the other hand, a kinetic approach has just permitted a relatively complete investigation of the properties of this system. We shall therefore analyze in detail the behaviour of a free-electron gas in a radiation field*.

* Footnote see next page.

We consider a low-density plasma in a field of intense, low-frequency radiation. We restrict ourselves from the outset to the regime of non-relativistic electron energies.

Any interaction of radiation with the heavy nuclei is obviously unimportant as compared with direct interaction with electrons. Clearly, the condition of electrical neutrality of the plasma cannot be violated: every change in the distribution of electron density induces a corresponding displacement of nuclei. If, however, the motion of the electrons is caused by a radiation field with a frequency that is large compared with the Langmuir frequency of the plasma, ω_0 , then the collective motion of nuclei and electrons turns out to be insignificant. (More precisely, it can be shown that the frequencies ω of the radiation field must be greater than $\omega_0 c/u$, where u is the thermal velocity of electrons.) With motion in this type of field, the electrons acquire a certain independence of the nuclei, and interact with the radiation as free particles. It is therefore sufficient for our purposes to consider an idealised system — a low-density electron gas in a radiation field. We shall ignore the more subtle effects associated with the collective motions in the plasma.

In examining the behaviour of free electrons in a low-frequency radiation field, two approaches are available. First, it is possible to perform the classical calculation for the motion of individual electrons in a wave field, and then to consider the statistical behaviour of a gas composed of these electrons. The second possibility is to discuss from the outset the behaviour of a system consisting of a free-electron gas and a photon gas. The properties of this system in the classical approximation can be obtained by taking the classical limit $\hbar \rightarrow 0$.

The classical analysis is possible and valid in a region of sufficiently low frequencies and high radiation densities. Here the occupation numbers $n(\mathbf{k})$ of photons (cf. §101 of Part V) are large, and, as usual for large quantum numbers, the classical limit can be taken in the quantum mechanical formulae (see §100 of Part V).

It should be emphasised, first of all, that the motion of an electron in a spectral radiation field (composed of a continuous set of frequencies and wave vectors) is essentially different from its motion in the field of an individual monochromatic wave. To illustrate this, consider the simple example of motion of an electron in a field of two waves which have the same frequen-

* The contents of this chapter are based on the following papers: J. Zeldowich and E. Levich, Soviet Physics JETP 55 (1968) 2423; E. Levich, Soviet Physics JETP 60 (1971) 112; T. Peraud, J. de Phys. 29 (1968) 88, 306, 872; A. Kompaneets, Soviet Phys. JETP 91 (1960) 876; T. Zeldowich and E. Levich, Soviet Physics JETP Letters 11 (1970) 57.

cy but differ in phase. The equation of motion of a non-relativistic electron can be written in the form

$$m \frac{d\mathbf{v}}{dt} = e \{ \mathbf{E}_1 + \mathbf{E}_2 + \frac{1}{c} [\mathbf{v} \times (\mathbf{H}_1 + \mathbf{H}_2)] \} \quad (70.1)$$

We solve (70.1) by successive approximations; that is, we put

$$\begin{aligned} \mathbf{v} &= \mathbf{v}_0 + \mathbf{v}_1, & |\mathbf{v}_1| &\ll |\mathbf{v}_0|, \\ m \frac{d\mathbf{v}_0}{dt} &\simeq e(\mathbf{E}_1 + \mathbf{E}_2), & m \frac{d\mathbf{v}_1}{dt} &\simeq \frac{e}{c} [\mathbf{v}_0 \times (\mathbf{H}_1 + \mathbf{H}_2)]. \end{aligned} \quad (70.2)$$

If we represent the field as

$$\mathbf{E}_1 = \mathbf{E}_0 e^{i\omega t} + \text{complex conjugate}, \quad (70.3)$$

$$\mathbf{E}_2 = \mathbf{E}_0 e^{i\omega t + i\alpha} + \text{complex conjugate}, \quad (70.4)$$

$$\mathbf{H}_1 = \mathbf{E}_1 \times \mathbf{n}_1, \quad (70.5)$$

$$\mathbf{H}_2 = \mathbf{E}_2 \times \mathbf{n}_1, \quad (70.6)$$

we immediately obtain

$$\mathbf{v}_0 = \frac{e(\mathbf{E}_1 + \mathbf{E}_2)}{i\omega m} - \text{complex conjugate}. \quad (70.7)$$

For the average (over a period) $\overline{d\mathbf{v}_1/dt}^T$ we have

$$\begin{aligned} \overline{\frac{d\mathbf{v}_1}{dt}}^T &= \frac{e^2}{i\omega cm^2} \overline{\{ \mathbf{E}_1 \times \mathbf{H}_2 + \mathbf{E}_2 \times \mathbf{H}_1 \}}^T = \\ &= \frac{2e^2 E^2}{\omega cm^2} \sin \alpha. \end{aligned} \quad (70.8)$$

It is clear that if the phase shift α is non-zero, the electron experiences a mean constant (systematic) force $2e^2 E_0^2 / \omega cm^2$. In contrast with the periodic motion in the field of a single wave, the electron now executes both oscillations and systematic motion. Thus an electron systematically accumulates energy. The

electron performs strictly periodic motion only when the two waves are parallel, or when the total magnetic field ($\mathbf{H}_1 + \mathbf{H}_2$) has a phase shift of $\frac{1}{2}\pi$ relative to the total electric field ($\mathbf{E}_1 + \mathbf{E}_2$). It should be stressed that the electron interacts simultaneously with the two waves, and therefore the constant force is proportional to the product of the field intensities.

If we now consider the real spectral field, composed of an infinite set of harmonics ω_i , it is apparent that the motion of an electron in a radiation field becomes a statistical problem. The only significant qualities are averages over all values of random phases. Thus, even the motion of a single electron in a spectral electromagnetic field requires statistical treatment. The classical approach necessitates the superposition of an infinite number of frequencies, followed by averaging with respect to random phases. This is fairly cumbersome. Therefore we turn to the second of the above approaches and consider from the very beginning the behaviour of the quantum statistical system (free electron gas and photon gas).

In a system consisting of free electrons and photons there are two fundamental interaction processes: (1) Compton scattering, and (2) free-free transitions (bremsstrahlung or absorption of photons in collisions between electrons). This system (electron gas + photon gas) also involves radiation-free collisions between electrons. In a real plasma these processes are supplemented by collisions of electrons with nuclei and the corresponding free-free radiation transitions, the latter being more significant than free-free transitions in the case of electron collisions.

The Compton scattering of a photon by a free electron is accompanied by changes both in the wave vector and in the photon frequencies. It is simplest of all to describe the change of these quantities in a reference frame K_0 in which the electron was initially at rest. In this reference frame we can express the change of momentum of an electron as

$$\Delta \mathbf{p}_0 = \frac{\hbar \omega_0}{c} (\mathbf{l} - \mathbf{l}') + \frac{\hbar}{c} (\omega_0 - \omega'_0) \mathbf{l}' = \Delta \mathbf{p}_1 + \Delta \mathbf{p}_2, \quad (70.9)$$

where

$$\Delta \mathbf{p}_1 = \frac{\hbar \omega_0}{c} (\mathbf{l} - \mathbf{l}'); \quad (70.10)$$

\mathbf{l} and \mathbf{l}' are respectively unit vectors in the directions of propagation of incident and scattered photons, (ω_0 and ω'_0 are corresponding frequencies). The quantity $\Delta \mathbf{p}_1$ represents the change in momentum of an electron due to scat-

tering without a change in photon frequency, i.e., due to a simple change in its direction of motion. The quantity $\Delta \mathbf{p}_2$ defines the change in momentum of an electron due to transfer to it of energy from a photon in the act of inelastic scattering. Using the expression for the change in frequency in the Compton effect (cf. eq. (17.11) of Part II), we can write

$$\omega_0 - \omega'_0 = \Delta\omega_0 = \frac{\hbar\omega_0^2}{mc^2} (1 - \mathbf{l} \cdot \mathbf{l}') \quad (70.11)$$

from which it follows that

$$\Delta \mathbf{p}_2 = \frac{\hbar^2 \omega_0^2}{mc^3} \mathbf{l}' (1 - \mathbf{l} \cdot \mathbf{l}') . \quad (70.12)$$

Obviously the ratio

$$\frac{|\Delta \mathbf{p}_2|}{|\Delta \mathbf{p}_1|} \simeq \frac{\hbar\omega_0}{mc^2} \ll 1$$

is very small in the non-relativistic approximation.

We note further that in a laboratory reference frame the change in frequency of a photon due to scattering by an electron with momentum \mathbf{p} has the form

$$\begin{aligned} \Delta\omega &\simeq \Delta\omega_0 + \omega_0 \left(1 + \frac{\mathbf{p} \cdot \mathbf{l}}{mc}\right) - \omega_0 \left(1 + \frac{\mathbf{p} \cdot \mathbf{l}'}{mc}\right) \simeq \\ &\simeq \Delta\omega_0 + \frac{\omega_0}{mc} (\mathbf{p} \cdot (\mathbf{l} - \mathbf{l}')) \simeq \frac{\hbar\omega_0^2}{mc^2} (1 - \mathbf{l} \cdot \mathbf{l}') + \frac{\omega}{mc} (\mathbf{p} \cdot (\mathbf{l} - \mathbf{l}')) . \end{aligned} \quad (70.13)$$

In the same non-relativistic approximation, we can write $\omega_0 \simeq \omega$. The probability of Compton scattering of a photon with wave vector variation $\mathbf{k} \rightarrow \mathbf{k}'$ in the solid angle $d\Omega$ can be written as

$$d\omega_{\mathbf{k} \rightarrow \mathbf{k}'} = cN_e [1 + n(\mathbf{k}')] n(\mathbf{k}) d\sigma , \quad (70.14)$$

where $n(\mathbf{k})$ is the number of photons with wave vector \mathbf{k} and N_e is the number of electrons per unit volume. The quantity $[1 + n(\mathbf{k})]$ (cf. eq. (103.7) of Part V) is connected with the Bose statistics of photons. In the long-wave approximation, which is relevant for our purposes, the effective cross section

$d\sigma$ in the rest-frame K_0 is given by the Thomson formula (cf. §36 of Part I)

$$d\sigma = \sigma d\Omega = \left(\frac{e^2}{mc^2} \right)^2 \frac{1}{2} (1 + \cos^2 \alpha) d\Omega, \quad (70.15)$$

where α is the angle of scattering.

We see that it is possible to distinguish between two types of scattering: spontaneous, proportional to the number of photons $n(\mathbf{k})$, and induced, proportional to $n(\mathbf{k})n(\mathbf{k}')$. In the classical limit of large occupation numbers the probability of induced scattering is much greater than that of spontaneous scattering. This is a general relationship between spontaneous and induced processes. The latter always correspond to the classical behaviour of a system. It follows that, in the classical limit discussed above, the mean motion of an electron in a spectral radiation field can be found by considering the induced scattering of photons by the electron.

Bearing in mind the classical limit $\hbar \rightarrow 0$, we replace the number of photons corresponding to a given vector by their spectral distribution $\rho(\omega, \mathbf{l})$. In accordance with eq. (76.9) of Part III, this spectral distribution can be related to the number of photons by the formula

$$\rho(\omega, \mathbf{l}) = \frac{\hbar \omega^3}{\pi^2 c^3} n(\omega, \mathbf{l}). \quad (70.16)$$

The spectral distribution determines the mean radiation energy at a given frequency. Obviously $\rho(\omega, \mathbf{l})$ is independent of the value of Planck's constant \hbar . We see that in the classical limit

$$dw \simeq c N_e \frac{\pi^4 c^6}{\hbar^2 \omega^6} \rho(\omega, \mathbf{l}) \rho(\omega, \mathbf{l}') d\sigma. \quad (70.17)$$

It should be emphasised that the number of photons remains unchanged in the scattering process.

Free-free bremsstrahlung represents the radiation or absorption of photons by an electron which is moving in the field of a nucleus or of another electron. The probability W_{ff} of a bremsstrahlung process is proportional to the square of the density N_e of electrons, and has the form*

$$W_{ff} \sim N_e^2 / \omega^3. \quad (70.18)$$

* See, for example, B. Heitler, *Quantum theory of radiation* (Clarendon Press, Oxford, 1954).

The probability of bremsstrahlung processes in the case of a low-density electron gas and high radiation density is, in general, much smaller than the probability of induced Compton scattering. For this reason we shall neglect bremsstrahlung except at low frequencies. Since W_{ff} increases very rapidly with decreasing frequency, bremsstrahlung processes become predominant when the frequency is sufficiently small. Hence we have a second limitation on the frequency range under consideration, namely

$$\omega > \omega_*, \quad (70.19)$$

where ω_* is the frequency at which (for fixed densities of the electron and photon gases) the change in the photon distribution function due to bremsstrahlung becomes greater than that due to the Compton effect. The frequency ω_* clearly decreases with increasing density N_e and increases with increasing $\rho(\omega, I)$.

We have assumed so far that the low-density plasma interacts with radiation of a known spectral distribution $\rho(\omega, I)$. It is however possible to formulate the problem in a different way, namely, the change of the spectral radiation distribution function due to the interaction with the electron gas, which also arises. For example, in the case of interaction with an electron gas which is in an equilibrium state, the spectral radiation distribution changes (evolves) with time, while the state of the electrons remains practically invariant. This process will also be discussed below, with the same approximations as before.

§71. Kinetic equations for electrons and photons

First of all we consider the formulation of the kinetic equations for electrons and photons. In the kinetic equation

$$\frac{df}{dt} = J, \quad (71.1)$$

where $f = f(\mathbf{p}, t)$ is the distribution function for electrons, the collision integral can be expressed as

$$J = J_c + I_{ee}. \quad (71.2)$$

Here I_{ee} is the collision integral for collisions between electrons, and is given by the Landau formula (34.7).

The change in distribution function of electrons due to Compton scattering is determined by the integral I_C . In what follows we shall assume that the system is homogeneous in space. Thus we shall suppose the distribution function $f(\mathbf{p}, t)$ for electrons to depend on the momentum vector and the time. Correspondingly, the distribution function for photons, i.e. their occupation numbers, will be presented as $n = n(\mathbf{k}, t)$. Taking this into account as well as formula (70.15) for the scattering probability, we can write the collision integral for electrons with photons J_C in the form

$$J_C = - \int dW_{\mathbf{k} \rightarrow \mathbf{k}'} f(\mathbf{p}, t) + \int dW_{\mathbf{k}' \rightarrow \mathbf{k}} f(\mathbf{p}', t), \quad (71.3)$$

where

$$dW_{\mathbf{k}' \rightarrow \mathbf{k}} = c [1 + n(\mathbf{k}, t)] n(\mathbf{k}', t) d\sigma d\mathbf{k}' = d\omega_{\mathbf{k}' \rightarrow \mathbf{k}} / N_e. \quad (71.3')$$

The first integral in (71.3) expresses the reduction in the number of electrons having momentum \mathbf{p} , while the second represents the arrival of electrons in this state as a result of collisions with photons.

For the case in which we are interested, that is of low radiation frequencies, the change in momentum of an electron through a single collision with a photon is small compared with its mean value. Therefore, the distribution function is a slowly-varying function of its argument, and, in accordance with results of § 10, the kinetic equation reduces to an equation of the Fokker-Planck type. Without repeating the calculations of § 10, we can write down the Fokker-Planck equation for three-dimensional momentum space, that is, for $\lambda = p_x, p_y, p_z$, in the form

$$\frac{df}{dt} = - \frac{\partial}{\partial p_i} \left\{ a_i f - \frac{\partial}{\partial p_k} (D_{ik} f) \right\} + I_{ee}, \quad i, k = x, y, z, \quad (71.4)$$

where the diffusion coefficient D_{ik} and the mobility a_i in momentum space are given by the formulae

$$D_{ik} = \int \Delta p_i \Delta p_k dW_{\mathbf{k} \rightarrow \mathbf{k}'}, \quad (71.5)$$

$$a_i = \int \Delta p_i dW_{\mathbf{k} \rightarrow \mathbf{k}'}, \quad (71.6)$$

Substituting for $dW_{\mathbf{k} \rightarrow \mathbf{k}'}$ from (71.3') we obtain

$$D_{ik} = \langle \Delta p_i \Delta p_k \rangle = \int \Delta p_i \Delta p_k n(\mathbf{k}) [1 + n(\mathbf{k}')] c d\mathbf{k} d\sigma, \quad (71.7)$$

$$a_i = \langle \Delta p_i \rangle = \int \Delta p_i n(\mathbf{k}) [1 + n(\mathbf{k}')] c \, d\mathbf{k} \, d\sigma, \quad (71.8)$$

where $\langle \rangle$ denotes the average over the photon distribution function. Note that the diffusion coefficients in momentum space characterise the random accumulation of energy by electrons, while the mobility characterises the systematic force acting on an electron. In later sections we shall calculate the kinetic coefficients (71.7) and (71.8), and find the solution of eq. (71.3) under various conditions.

It turns out to be expedient to calculate the kinetic coefficients in the reference frame K_0 in which the electron is at rest, since in this frame the scattering cross section has a particularly simple form. Moreover, as we have noted earlier, there is a simple expression for the momentum transferred.

Before proceeding to discussing the kinetic equation for electrons, we obtain the corresponding kinetic equation for photons. In this case we restrict ourselves to an isotropic distribution of photons, i.e. $n = n(\omega, t)$.

Obviously

$$\frac{dn}{dt} = J_c^{(n)}, \quad (71.9)$$

where $J_c^{(n)}$ is the collision integral for collisions between photons and free electrons:

$$\begin{aligned} J_c^{(n)} = & - \int n(\omega, t) [1 + n(\omega', t)] c \, d\sigma f(\mathbf{p}, t) \, d\mathbf{p} + \\ & + \int n(\omega', t) [1 + n(\omega, t)] c \, d\sigma f(\mathbf{p}', t) \, d\mathbf{p}. \end{aligned} \quad (71.10)$$

The first term gives the decrease in the number of photons of frequency ω due to collisions with an electron of momentum \mathbf{p} , while the second term gives the arrival of photons in this state.

In the equilibrium state $dn/dt = df/dt = 0$. As we saw earlier, the integral I_{ee} for collisions of electrons among themselves vanishes for a Maxwell distribution. It can easily be verified that the integrals J_c and $J_c^{(n)}$ vanish in the case when electrons have a Maxwell distribution with temperature T and a photon distribution as given by

$$n(\mu, \omega) = \frac{1}{e^{(\hbar\omega - \mu)/kT} - 1}, \quad (71.11)$$

i.e. a Bose distribution function with the same temperature T and chemical

potential μ (not, in general, equal to zero). This last result allows a simple interpretation: in our approximation, when radiation and absorption of photons are neglected, their total number in the system is fixed. We recall (cf. eq. (76.6) of Part III) that, in general, when the number of photons is not fixed, the value $\mu = 0$ corresponds to a minimum of free energy. It is clear, therefore, that the processes of absorption and radiation, which occur during a sufficiently long period of time, transform (71.11) into the equilibrium Planck distribution and $\mu \rightarrow 0$.

Passing now to a consideration of the kinetic equation for photons (71.9), we shall show that it can also be reduced to a differential equation of the Fokker-Planck type. In the case of non-relativistic electrons and sufficiently soft radiation, $\hbar\omega \ll mc^2$. Thus formula (70.11) shows that the change in frequency of a photon through scattering is small. We suppose that the electrons can be characterised by an equilibrium Maxwell distribution with temperature T . Then, introducing a new variable $x = \hbar\omega/kT$, such that the quantity $\hbar\Delta\omega/kT$ is small, we can expand the photon distribution function in a series of powers of Δx . Restricting ourselves to the first terms of this expansion we now obtain

$$\begin{aligned} J_c^{(1)} = & - \int d\mathbf{p} [1 + n(x + \Delta x)] n(x) c d\sigma f(\epsilon) + \\ & + \int d\mathbf{p} n(x + \Delta x) [1 + n(x)] c d\sigma f(\epsilon) = [1 + n(x)] \int d\mathbf{p} f(\epsilon) \Delta x c d\sigma + \\ & + \frac{1}{2} \left[\frac{\partial^2 n}{\partial x^2} + 2(1 + n) \frac{\partial n}{\partial x} + n(1 + n) \right] c \int d\mathbf{p} d\sigma (\Delta x)^2 f(\epsilon) = J_c^{(1)} + J_c^{(2)}. \end{aligned} \quad (71.12)$$

The evaluation of the second integral in (71.12) is easy enough:

$$J_c^{(2)} = \frac{1}{2} \left[\frac{\partial^2 n}{\partial x^2} + 2(1 + n) \frac{\partial n}{\partial x} + n(1 + n) \right] c I,$$

where

$$I = c \int d\mathbf{p} d\sigma (\Delta x)^2 f(\epsilon) = \sigma_T N_e \left(\frac{kT}{mc^2} \right) x^2. \quad (71.13)$$

Here, in the expression for $\Delta x = \hbar\Delta\omega/kT$, we have used (70.13), and have denoted by σ_T the total Thomson cross section, $\sigma_T = \frac{8}{3}\pi(e^2/mc^2)^2$.

The evaluation of the first integral in (71.12) is much more difficult. How-

ever, this calculation can be replaced by the following general argument. The law of conservation of the number of particles in the process of scattering – in our case, the number of photons – requires that the kinetic equation should have the form

$$\frac{\partial n}{\partial t} = J_c^{(f)} = -\nabla \cdot \mathbf{j}, \quad (71.14)$$

where \mathbf{j} is the photon flux in momentum space. Because of the isotropy of the photon distribution function, the divergence of the flux has the form

$$\nabla \cdot \mathbf{j} = \frac{1}{x^2} \frac{\partial}{\partial x} \{x^2 j(x)\}. \quad (71.15)$$

There must be a linear relationship between the flux j and $\partial n / \partial x$. Moreover in the equilibrium state (for a Planck distribution) the flux $j(x)$ should be equal to zero. Thus we may write

$$j(x) = \left[\frac{\partial n}{\partial x} + A(x) \right] g(x),$$

where $g(x)$ and $A(x)$ – two unknown functions – are to be found. It is easy to verify that for a Planck distribution the equality

$$(1+n)n = -\frac{\partial n}{\partial x}, \quad (71.16)$$

holds: it is easily verified by a direct calculation. Hence both of the above requirements on j will be satisfied if we write $j(x)$ in the form

$$j(x) = \left[\frac{\partial n}{\partial x} + n(1+n) \right] g(x), \quad (71.17)$$

Correspondingly we find for $J_c^{(f)}$ the general expression

$$J_c^{(f)} = -\frac{1}{x^2} \frac{\partial}{\partial x} \left\{ x^2 g(x) \left[\frac{\partial n}{\partial x} + n(1+n) \right] \right\} \quad (71.18)$$

In order that the two expressions for $J_c^{(1)}$, i.e. (71.18) and (71.12), should coincide, it is necessary to put $g(x) = -x^2 \sigma_T N_e kT/mc$. Thus we have finally

$$\begin{aligned} \frac{\partial n(x, t)}{\partial t} = \frac{kT}{mc^2} (\sigma N_e c) \frac{1}{x^2} \frac{\partial}{\partial x} x^4 \times \\ \times \left\{ \frac{\partial n(x, t)}{\partial x} + n(x, t) + [n(x, t)]^2 \right\}. \end{aligned} \quad (71.19)$$

Notice that although (71.19) is similar to a Fokker–Planck equation, it is not quite the usual equation of this type. The presence of the factor $1 + n$ in the transition probability has resulted in the fact that (71.19) is nonlinear with respect to the unknown function $n(x)$.

The kinetic equations for electrons, (71.4), and photons, (71.19), obviously constitute a coupled system. To find the general solution of this system is an extremely complex problem, so that we shall consider only two limiting cases: (1) to find the electron distribution $f(\mathbf{p}, t)$ for an arbitrary, prescribed, photon distribution; (2) to find the photon distribution function $n(\omega, t)$ for a given equilibrium distribution of electrons.

§ 72. Kinetics of Bose condensation in a photon gas

We shall first consider the second of the above problems: the investigation of the change in the properties of non-equilibrium radiation as a result of interaction with free electrons. In other words, we shall find the law of evolution in time of a photon distribution $n(\omega, t)$ due to interaction with an electron gas. As we remarked earlier, it will be assumed here that the state of the electron gas is prescribed and stationary. Namely, we shall suppose that the electron gas is in an equilibrium state for all time, and is characterised by some temperature T_e .

We shall completely neglect free–free transition processes, and assume that scattering is the only form of interaction between photons and electrons. The role of absorption will be considered later.

The kinetic equation (71.19), which described the time variation of $n(\omega, t)$, must be supplemented by initial conditions. Let $n_p(\omega)$ be the equilibrium (Planck) distribution of photons corresponding to the temperature T_e . Two cases are then possible:

(1) at the initial moment, $t = 0$, $n(\omega, 0)$ is such that the total number of photons in the system

$$N \sim \int_0^{\infty} n(x, 0) x^2 dx$$

is larger than the total number of photons in equilibrium radiation

$$N_P \simeq \int_0^{\infty} n_P(x) x^2 dx, \quad (72.1)$$

that is, $N > N_P$;

(2) at the initial instant

$$N < N_P. \quad (72.2)$$

It is clear that the evolution of $n(x, t)$ will be different in the two cases. In the first case the photons will tend to lose energy in collisions with electrons, and their spectral evolution will, on average, consist of a trend downwards, i.e. towards the energy axis ('cooling' of the photon gas). In the other case the photons will in the mean gain energy through collisions ('heating' of photons).

Since we neglect the absorption of photons, their total number will remain constant during the evolution process. We begin by considering the 'cooling' of the photon gas. As was pointed out earlier, we are interested here in the region of comparatively low frequencies. In this region the exchange of energy between a photon and an electron during the process of scattering is small, and the kinetic equation (71.19) is valid. Moreover, the occupation numbers $n(x, t)$ will be large. As was pointed out earlier, this is the region where classical electrodynamics holds.

If we assume that $n \gg 1$, and, furthermore, that the inequality

$$n^2 \gg \left| \frac{\partial n}{\partial x} \right|, \quad (72.3)$$

is fulfilled, then eq. (71.19) can be substantially simplified, i.e. we can omit the small (compared with n^2) terms n and $\partial n / \partial x$. The validity of assumption (72.3) can be verified *a posteriori* by direct calculation. Thus, we have in-

stead of (71.19)

$$\frac{\partial n}{\partial t'} \simeq \frac{1}{x^2} \frac{\partial}{\partial x} (x^4 n^2), \quad (72.4)$$

where we write $t' = (\sigma N_e c k T / mc^2) t$. Introducing a new unknown function $f = x^2 n(x, t)$, we find

$$\frac{\partial f}{\partial t'} = \frac{\partial f^2}{\partial x}. \quad (72.5)$$

The solution of this equation in characteristics is

$$x = F(f) - 2ft', \quad (72.6)$$

where the form of the function $F(f)$ is determined by the initial conditions. The meaning of this result is best understood by reference to the example depicted in fig.VI.14. According to (72.6), all points located on the initial curve $f(x, 0) = x^2 n(x, 0)$ (curve 1) move along characteristics, which are straight lines parallel to the axis in the direction of decreasing x (curve 2 in fig.VI.14), and with velocity proportional to x . The time required for a given point to reach the f -axis is obviously given by the expression

$$\tau = F(f)/2f. \quad (72.7)$$

The solution (72.6) is formally applicable for all values of x , both positive and negative. In the course of time, therefore, $f(x, t)$ must take the form re-

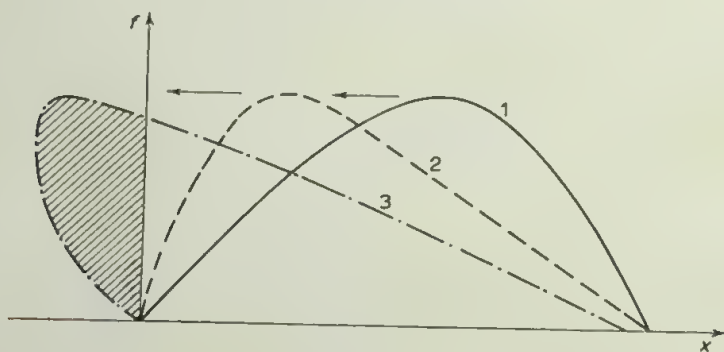


Fig. VI.14

presented by curve 3 of fig.VI.14. It is clear, however, that since the variable x represents the energy of the photons, it is impossible for x to attain negative values. The particles which have arrived at the state $x = 0$, i.e. at the zero-energy state, cease to move. In other words, the photons accumulate in a state with zero energy, which means in a Bose condensation state.

The number of particles passing to this state is formally determined by the shaded area in fig.VI.14. It must be emphasised, however, that this picture of the kinetics of photon transition to a zero-energy state is incomplete, and must not be accepted literally. In a very low frequency region absorption always plays an important role (cf. §70), no matter how small the concentration of the electron gas (or, more precisely, of the plasma).

If the system (electron gas + photon gas) is closed for a sufficiently long time, it will always evolve to a statistical equilibrium and the photon distribution function will become Planckian. Therefore it would be senseless to look for the stationary solutions of eq. (71.19) without taking bremsstrahlung into account. On the other hand, formula (72.4) gives the dynamics of the change in the photon spectrum for the time t^* and in that frequency range where bremsstrahlung is of no importance. The time-variation of the initial distribution $n(x, 0)$ is illustrated by the curves drawn in fig.VI.14. Each of these is a kind of snapshot of the distribution for t lying in the interval $0 < t \leq t^*$. Since the kinetics of the process depends essentially on the form of the initial distribution, it is interesting to examine the case when $f(x, 0)$ has the form shown in fig. VI.15. According to (68.6), the function $f(x, t)$ will deform in the course of time as shown in fig.VI.15. The rate of approach to the ordinate axis will be the greater, the greater the value of f . The upper parts of the curve are being advanced while the lower ones are being retarded. As a result, the distribution function can be deformed to such an extent that the curve of $f(x, t)$ turns out not to be single valued — this is illustrated in fig.VI.15.

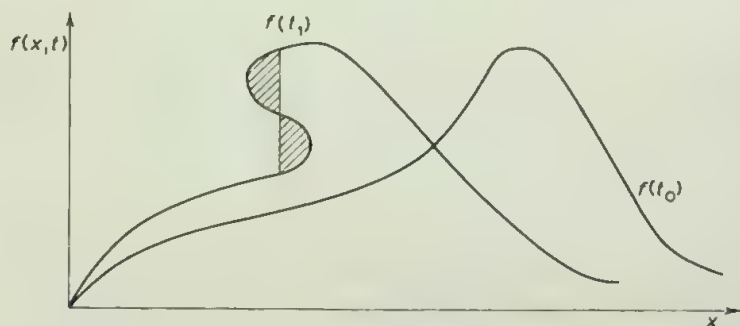


Fig. VI.15

A similar situation arises in the case of the generation of shock waves in the uniform flow of an inviscid fluid. In reality, of course, $f(x, t)$ does not become multivalued. Near the shock front the derivative $\partial n / \partial x$ increases without limit; inequality (72.5) is violated and eq. (72.4) becomes inapplicable. Moreover, the change in the distribution function becomes abrupt in the vicinity of the front, and the general equation (71.19) is also inapplicable. A detailed picture of the motion of the front $f(x, t)$ in this case becomes much more complicated.

The second stage of evolution begins when $t > t^*$. Due to absorption, an equilibrium Planck distribution is established in the photon system, with the temperature of the photons equal to the temperature T_e of the electrons.

Let us now consider the second case, where inequality (72.2) is satisfied at the initial instant. The evolution of the photon spectrum again takes place in two stages. In the first stage, with $t < t^*$, a Bose-Einstein distribution with non-zero chemical potential is set up in the photon system. Here the mean energy of photons becomes equal to the mean energy of electrons.

Next, absorption and radiation processes come into play. The chemical potential begins to vary with time, and after the elapse of time $t \gg t^*$, $\mu \rightarrow 0$, i.e. a Planck distribution is established in the system.

§73. Mobility of an electron in a radiation field

Before proceeding to consider the behaviour of a system of electrons in a radiation field, we must deal with the important question of the behaviour of a single free electron in such a field.

The behaviour of an electron in the field of an electromagnetic wave was discussed in §29 of Part I. There we saw that the electron scatters an electromagnetic wave impinging on it. It is clear, however, that, since the process of scattering involves transfer of momentum from the wave to the electron, some mean force is generated which acts on a particle radiating secondary scattered waves. By eq. (29.2) of Part I, the mean force (over a period) acting on a radiating particle (in a reference frame in which the particle is at rest: henceforth we shall omit the subscript zero for this reference frame, so long as no ambiguity arises from doing so) is

$$\begin{aligned} \mathbf{F}_0 &= \frac{2}{3} \frac{e^2}{c^3} \frac{d^2 \mathbf{v}}{dt^2} \simeq \frac{2e^2}{3c^3} \left\{ \frac{e \dot{\mathbf{E}}}{m} + \frac{e}{mc} (\dot{\mathbf{v}} \times \mathbf{H}) \right\} \simeq \\ &\simeq \frac{2e^4}{3m^2 c^4} (\overline{\mathbf{E} \times \mathbf{H}}) = \frac{8}{3} \pi \left(\frac{e^2}{mc^2} \right)^2 \frac{\overline{\mathbf{E}^2}}{4\pi} \mathbf{l}, \end{aligned}$$

where \mathbf{l} is unit vector in the direction of propagation of the wave. Introducing the effective Thomson scattering cross section and the energy density of radiation u_0 , we arrive at the formula for radiation pressure force

$$\mathbf{F}_0 = \sigma_T u_0 \mathbf{l} = \frac{\sigma_T}{c} \mathbf{q}, \quad (73.1)$$

where \mathbf{q} is the energy flux in a single wave. The force acting on the electron — the radiation reaction force — is independent of the frequency. Therefore, the force acting on the electron due to a spectral field generated by a collection of waves with different frequencies can also be expressed in the form

$$\mathbf{F}_0 = \frac{\sigma_T}{c} \mathbf{q}. \quad (73.2)$$

Here \mathbf{q} denotes the total energy flux in the radiation field

$$\mathbf{q} = \int c \hbar \omega n(\omega, \mathbf{l}) \mathbf{l} d\mathbf{k} = \pi^2 c^4 \int \rho(\omega, \mathbf{l}) \frac{1}{\omega^2} d\mathbf{k}. \quad (73.3)$$

In place of the energy density we have introduced the spectral function $\rho(\omega, \mathbf{l})$.

In the transition from (73.1) to (73.2), i.e. from the force on an electron due to a monochromatic wave to the equivalent force in a spectral radiation field it has been implicitly assumed that additivity of the forces due to the individual waves is valid. In this summation the effect of induced scattering, mentioned previously, has been ignored. Here the following fact turns out to be significant: if we do not take into account that part of the change in momentum, Δp_2 , which is associated with the change in frequency at scattering, and whose contribution to the total change in momentum we know to be small (cf. (70.9)), and if we also ignore any change in the spectral form of the distribution function due to change in frequency, then the contribution of the induced scattering to the reaction is identically zero.

On the other hand, the preceding discussion makes it clear that if we do take into account frequency variation the formula for the reaction is essentially different.

Thus, we shall have to find an exact expression for the reaction without any arbitrary assumptions. By an 'exact' expression we mean an expression for the reaction in the classical non-relativistic approximation (without quantum or relativistic corrections). For this calculation it is convenient (although

in principle not obligatory) to use a quantum mechanical approach, based on the application of the general formula for mobility a_i . Going to the classical limit $\hbar \rightarrow 0$ in the resultant formulae we obtain the classical expression for the mean force. This example illustrates the usefulness of the technique discussed in the previous section: finding classical formulae from quantum formulae in the limit $\hbar \rightarrow 0$.

In accordance with (71.8) the mean force or the mobility is given by the general formula

$$\langle \Delta \mathbf{p} \rangle = \int c(\Delta \mathbf{p}_1 + \Delta \mathbf{p}_2) n(\mathbf{k}) [1 + n(\mathbf{k}')] d\sigma d\mathbf{k}. \quad (73.4)$$

We now use the fact that the change in frequency $\Delta\omega$ of a photon for a single scattering at an electron is small. (It is convenient here to use the frame K_0 ; the subscript zero is omitted for simplicity.) We write

$$n(\mathbf{k}) \equiv n(\omega, \mathbf{l}); \quad n(\mathbf{k}') \equiv n(\omega', \mathbf{l}')$$

where we have made the approximation

$$n(\omega', \mathbf{l}') \simeq n(\omega, \mathbf{l}') + \frac{\partial n}{\partial \omega} (\omega' - \omega) = n(\omega, \mathbf{l}') - \frac{\partial n}{\partial \omega} \Delta\omega.$$

Replacing the occupation number $n(\omega, \mathbf{l})$ by the spectral distribution $\rho(\omega, \mathbf{l})$ we have

$$\begin{aligned} \langle \Delta \mathbf{p} \rangle &= \int \left\{ \frac{\hbar\omega}{c} (\mathbf{l} - \mathbf{l}') + \frac{\hbar}{c} (\Delta\omega) \mathbf{l}' \right\} \frac{\pi^2 c^3}{\hbar \omega^3} \rho(\omega, \mathbf{l}) \times \\ &\quad \times \left\{ 1 + \frac{\pi^2 c^3 \rho(\omega, \mathbf{l}')}{\hbar \omega^3} - \frac{\pi^2 c^3}{\hbar} \Delta\omega \frac{\partial}{\partial \omega} \left[\frac{\rho(\omega, \mathbf{l}')}{\omega^3} \right] \right\} c d\sigma d\mathbf{k} = \\ &= \int \left\{ \frac{\hbar\omega}{c} (\mathbf{l} - \mathbf{l}') + \frac{\hbar^2 \omega^2}{mc^3} \mathbf{l}' (\mathbf{l} - \mathbf{l}') \right\} \frac{\pi^2 c^4}{\hbar \omega^3} \rho(\omega, \mathbf{l}) \left\{ 1 + \frac{\pi^2 c^3}{\hbar} \frac{\rho(\omega, \mathbf{l}')}{\omega^3} - \right. \\ &\quad \left. - \frac{\pi^2 c^3}{\hbar} \Delta\omega \frac{\partial}{\partial \omega} \left[\frac{\rho(\omega, \mathbf{l}')}{\omega^3} \right] \right\} \frac{2\omega^2 d\omega}{(2\pi c)^3} d\Omega_k d\Omega, \end{aligned} \quad (73.5)$$

where $d\Omega_k$ is an element of solid angle for the incident photon and where we have substituted the value (70.11) for $\Delta\omega$.

We now take the limit $\hbar \rightarrow 0$ in formula (73.5). To do this it is necessary to

retain all the terms which do not contain \hbar . The terms $\sim 1/\hbar$ cancel exactly after integration over angles. As a result we obtain for the total force acting on the electron

$$\begin{aligned}
 \langle \Delta \mathbf{p} \rangle &= \int \sigma \frac{\pi^2 c^3}{\omega^2} \rho(\omega, \mathbf{l}) (1 - \mathbf{l} \cdot \mathbf{l}') d\mathbf{k} d\Omega + \\
 &+ \int \sigma \frac{\pi^4 c}{m \omega^2} \rho(\omega, \mathbf{l}) \rho(\omega, \mathbf{l}') \mathbf{l}' (1 - \mathbf{l} \cdot \mathbf{l}') \frac{d\omega}{4\pi^3} d\Omega_{\mathbf{k}} d\Omega - \\
 &- \int \sigma \frac{\pi^4 c}{m} \rho(\omega, \mathbf{l}) \frac{\partial}{\partial \omega} \left[\frac{\rho(\omega, \mathbf{l})}{\omega^3} \right] (1 - \mathbf{l}') \omega^2 (1 - \mathbf{l} \cdot \mathbf{l}') \frac{d\omega}{4\pi^3} d\Omega_{\mathbf{k}} d\Omega = \\
 &= \mathbf{F}_{\text{sp}} + \mathbf{F}_{\text{ind}}.
 \end{aligned} \tag{73.6}$$

It should be emphasised that $\langle \Delta \mathbf{p} \rangle$ represents the mean force in the classical approximation. Therefore exactly the same expression could have been obtained by corresponding calculations with the relations of classical electrodynamics. At the same time it should be stressed that in the classical approximation we do not use an expansion in terms of the powers of some small parameters. In this sense (73.6) is an exact formula.

We examine separately each of the terms in (73.6). The first term can be written in the form

$$\begin{aligned}
 \mathbf{F}_{\text{sp}} &= \int \sigma \frac{\pi^2 c^3}{\omega^2} \rho(\omega, \mathbf{l}) (1 - \mathbf{l}') d\mathbf{k} d\Omega = \\
 &= \sigma_T \pi^2 c^3 \left(\int \rho(\omega, \mathbf{l}) \frac{d\omega}{c^3 4\pi^3} d\Omega_{\mathbf{k}} \right) \mathbf{l} = \mathbf{F}_0.
 \end{aligned} \tag{73.7}$$

We see that \mathbf{F}_{sp} is the force due to spontaneous scattering. It coincides with the radiation pressure \mathbf{F}_0 obtained above. It is evident from its structure that the second term

$$\begin{aligned}
 \mathbf{F}_{\text{ind}} &= \int \frac{\sigma c \pi}{4m} \frac{\rho(\omega, \mathbf{l}) \rho(\omega, \mathbf{l}')}{\omega^2} \mathbf{l}' (1 - \mathbf{l} \cdot \mathbf{l}') d\omega d\Omega_{\mathbf{k}} d\Omega - \\
 &- \int \frac{\sigma \pi c}{4m} \rho(\omega, \mathbf{l}) \frac{\partial}{\partial \omega} \left[\frac{\rho(\omega, \mathbf{l}')}{\omega^3} \right] (1 - \mathbf{l}') \omega^2 (1 - \mathbf{l} \cdot \mathbf{l}') d\omega d\Omega_{\mathbf{k}} d\Omega,
 \end{aligned} \tag{73.8}$$

represents the force due to induced scattering. This expression contains under

the integral sign the product of the spectral functions $\rho(\omega, \mathbf{l})$ and $\rho(\omega, \mathbf{l}')$ of incident and scattered radiation.

The character of the force F_{ind} depends on the form of the function $\rho(\omega, \mathbf{l})$. If in particular the radiation is strictly monochromatic, i.e. the radiation density is non-zero only for a fixed value of the frequency and wave vector, then $\rho(\omega, \mathbf{l}') = 0$, and the induced force vanishes identically.

If $\rho(\omega, \mathbf{l})$ is the equilibrium Planck distribution $\rho_P(\omega, T)$, we obtain, after substituting for $\rho(\omega, T)$ and performing some simple calculations,

$$F_{\text{ind}} \simeq \frac{kT}{mc^2} F_{\text{sp}}. \quad (73.9)$$

Since in the non-relativistic region the radiation temperature satisfies the inequality $kT \ll mc^2$, F_{ind} represents a small correction to the spontaneous force F_{sp} .

Conversely, in an arbitrary high-intensity radiation field F_{ind} can be comparable with and even substantially larger than F_{sp} . We shall consider the induced reaction force for some concrete forms of spectral distribution. Suppose, for example, we take as the radiating element the uniform surface of a disc of radius R . Since all points of the disc radiate in the same way, the function $\rho(\omega, \mathbf{l})$ can be written in the form

$$\rho(\omega, \mathbf{l}) = G(\omega) \varphi(\mathbf{l}). \quad (73.10)$$

where the first factor represents the frequency and the second the spatial distribution of radiation. The expression for F_{ind} takes on a particularly revealing form if we choose a simple function $G(\omega)$, for example,

$$G(\omega) = \begin{cases} a/\omega^\gamma, & \omega_0 \leq \omega < \infty, \\ 0, & \omega < \omega_0. \end{cases} \quad (73.11)$$

We may note that this spectral distribution is associated with certain astrophysical sources of intensive radiation. Substitution of (73.10) and (73.11) now gives

$$F_{\text{ind}} = \frac{a^2 c \pi \sigma}{4m} \int_{\omega_0}^{\infty} \frac{d\omega}{\omega^{2\gamma+2}} \int \varphi(\mathbf{l}) \varphi(\mathbf{l}') (1 - \mathbf{l} \cdot \mathbf{l}') d\Omega_{\mathbf{k}} d\Omega. \quad (73.12)$$

The second integral in (73.12) vanishes, since its angular principal part changes sign under the transformation $\mathbf{l} \rightarrow -\mathbf{l}'$.

For a disc at a large distance $r \gg R$ from the surface we can write

$$\varphi(\mathbf{l}) = \varphi(\beta, \psi) = \Theta(\cos \beta - \cos \beta_0) \simeq \Theta(\beta^2 - \beta_0^2), \quad (73.13)$$

where β and ψ are polar and azimuthal angles, Θ is a step function, β_0 is the angle of the cone, and $\beta_0 \approx r/R \ll 1$. After evaluating (73.12) we obtain for the force

$$\mathbf{F}_{\text{ind}} = \frac{3\sigma_T \pi^2 (\gamma - 1)^2 \beta_0^2}{8m\omega_0^3 (2\gamma + 1)c} q\mathbf{q}. \quad (73.14)$$

We see that $|\mathbf{F}_{\text{ind}}|$ turns out to be proportional to q^2 , and for large values of radiation flux (and small values of ω_0) it can be substantially larger than $|\mathbf{F}_{\text{sp}}|$. Note, however, that this ratio falls off rapidly with distance from the radiating surface (like $\beta_0^6 \approx (r/R)^6$).

A second interesting example is the reaction force in a highly anisotropic field. We take as its source two streams of radiation in opposite directions. Let one of these be coherent (monochromatic), and directed along the positive z -axis:

$$n_1(\omega) = I_1 \delta(\omega - \omega_0). \quad (73.15)$$

The second beam is directed along the negative z -axis and is characterised by some spectral distribution $n_2(\omega)$. Using the general formula (73.8) for the induced reaction, we arrive, after a few calculations, at an expression for the component $F_{\text{ind}}^{(z)}$:

$$F_{\text{ind}}^{(z)} \simeq \frac{h^2 \sigma_T}{mc^5} I_1 \omega^{\frac{1}{2}} \left\{ \frac{\partial}{\partial \omega} [n_2(\omega) \omega^{\frac{3}{2}}] \right\} \bigg|_{\omega=\omega_0}. \quad (73.16)$$

Formula (73.16) shows that $F_{\text{ind}}^{(z)}$ can have either positive or negative sign, depending on the value of the derivative

$$\left\{ \frac{\partial}{\partial \omega} [n_2(\omega) \omega^{\frac{3}{2}}] \right\} \bigg|_{\omega=\omega_0}.$$

The spontaneous force F_{sp} in this example is always in the direction of that beam which has the higher density, i.e. whose energy flux q is greater. Thus,

depending on the sign of the above quantity and on the relation between the absolute values of F_{ind} and F_{sp} , two situations can arise: either the resultant reaction is in the direction of the total energy of the two radiation beams, or it is in a direction opposite to the total radiation flux. This seemingly paradoxical result again emphasises the importance of the induced scattering effect.

In conclusion we obtain an expression for the force on a moving electron with $v \ll c$. The spontaneous force can be deduced from (73.6) by a Lorentz transformation for the frequency with $v \ll c$ (Doppler effect!). It gives

$$F_{\text{sp}} = - \frac{4\sigma_T u_0 \mathbf{p}}{3cm}. \quad (73.17)$$

It turns out that the induced force is given by the formula

$$F_{\text{ind}} = \frac{\sigma_T \pi^2}{m^2} \mathbf{p} \left[\frac{13}{2} \int \left[\frac{\rho(\omega)}{\omega} \right]^2 d\omega - \frac{14}{15} \int \left\{ \frac{\partial}{\partial \omega} \left[\frac{\rho(\omega)}{\omega^3} \right] \right\}^2 \omega^6 d\omega \right]. \quad (73.18)$$

§74. System of electrons in an arbitrary radiation field

We can now proceed to find the stationary, non-equilibrium distribution of a system of electrons located in a given radiation field. We shall assume that the field is isotropic. In this case we assume that the distribution function n depends only on the frequency, and not on the direction of the wave vector, i.e. $n = n(\omega)$. In this simple case it is convenient to analyse the principal assertions of the theory.

From (71.4) we find for the stationary distribution of electrons

$$- \frac{\partial}{\partial p_i} D_{ii} f + a_i f = j_i, \quad (74.1)$$

where j_i is defined by formula (34.8).

To solve (74.1) it is necessary to know, apart from the mobility which we found before, the diffusion coefficient $D_{ii} = D$. Like the mobility, the diffusion coefficient is in general a function of the momentum.

For our purposes the expression for the mobility a_i can be limited to the approximation

$$a_i = \langle \Delta p_i \rangle \simeq F_{\text{sp}}^{(i)}, \quad (74.2)$$

where $F_{sp}^{(i)}$ is given by (73.17). The reason for this is as follows: as we shall see later, the electrons in the radiation field may arrive at a stationary state with a very high temperature. As will be shown later, if the induced reaction force F_{ind} is of the same order of magnitude as the spontaneous force F_{sp} , the electrons will acquire a mean energy exceeding mc^2 . But, as has been emphasised from the outset, the present theory is restricted to the case of non-relativistic electrons.

In calculating the diffusion coefficient from formula (71.5) we can confine ourselves to taking into account the change in momentum due to perfectly elastic scattering alone. Inelastic scattering gives only a small quantum correction. It is convenient to calculate the diffusion coefficient in the rest-frame of an electron K_0 .

Here we have for D :

$$\begin{aligned} D = \langle \Delta p_i^2 \rangle &= \frac{1}{3} c \int \sigma n_0(\omega_0) [1 + n_0(\omega'_0)] (\Delta \mathbf{p})^2 d\mathbf{k}_0 d\Omega = \\ &= \frac{1}{3} c \int \sigma \left(\frac{\hbar \omega_0}{c} \right)^2 (1 - l')^2 n_0(\omega_0) [1 + n_0(\omega'_0)] d\mathbf{k}_0 d\Omega, \end{aligned} \quad (74.3)$$

where the subscript zero refers to the number of photons and their frequency in this coordinate system.

To go over to a laboratory reference frame it is necessary to perform a Lorentz transformation. It turns out, however, that in calculating the diffusion coefficient in the non-relativistic approximation we need take into account only terms of order of magnitude zero with respect to the quantity p/mc . This is connected with the fact that the diffusion coefficient in eq. (74.1) is multiplied by the derivative $\partial f / \partial p_i$. As a result, as we shall see below, if we include terms of order p/mc in a_i , we need include only terms independent of p in D . In this case $D(\partial f / \partial p_i)$ will automatically be of the same order of magnitude as $a_i f$.

In keeping with this we can put

$$n_0(\omega_0) = n(\omega) = n \left(\omega_0 + \omega_0 \frac{p \cos \theta}{mc} \right) \approx n(\omega_0) + \frac{\partial n}{\partial \omega_0} \omega_0 \frac{p}{mc} \cos \theta$$

and we obtain for D

$$\begin{aligned} D &= \int \frac{1}{3} c \left(\frac{\hbar \omega}{c} \right)^2 (\cos \theta - \cos \theta')^2 n(\omega) [1 + n(\omega)] \sigma d\Omega d\mathbf{k} = \\ &= \frac{4}{3} \pi \sigma_{TC} \int \left(\frac{\hbar \omega}{c} \right)^2 n(\omega) [1 + n(\omega)] \frac{\omega^2 d\omega}{(2\pi c)^3}. \end{aligned} \quad (74.4)$$

On substituting (74.2) and (74.4) into (74.1) we have finally

$$\left\{ \frac{4}{3} \pi \sigma_T c \int \left(\frac{\hbar \omega}{c} \right)^2 n [1 + n] \frac{\omega^2 d\omega}{(2\pi c)^3} \right\} \frac{\partial f}{\partial \mathbf{p}} + \left\{ \frac{16}{3} \frac{\sigma_T \pi}{cm} \int \hbar \omega n \frac{\omega^2 d\omega}{(2\pi c)^3} \right\} \mathbf{p} f + \mathbf{j} = 0. \quad (74.5)$$

We shall try to find the solution of (74.5) in the form of a Maxwell distribution with some effective temperature θ , i.e. we put

$$f \sim e^{-p^2/2m\theta}. \quad (74.6)$$

As was shown in §34, if we put the Maxwell distribution (with any temperature θ) into the integral j_i , the latter automatically cancels exactly. Therefore eq. (74.5) can be satisfied if we set

$$\theta = \frac{\int n(\omega) [1 + n(\omega)] (\hbar \omega)^2 \omega^2 d\omega}{4 \int n(\omega) \hbar \omega \omega^2 d\omega}. \quad (74.7)$$

We see that when the effective temperature is given by formula (74.7) the Maxwell distribution becomes the exact solution of the equation (74.1) even for the case of a stationary state of the system of electrons in a non-equilibrium radiation field.

§75. General discussion of the results and range of the applicability of the theory

It is now possible to sum up the results of our calculations. It is necessary first of all to emphasize the remarkable peculiarity of the induced pressure force: this systematic force arises as a result of the action of a stochastic electromagnetic field. A clearer understanding of the interrelationship between spontaneous and induced pressure forces of the radiation can be gained by the introduction of the so-called brightness temperature, the latter being defined by the formula

$$kT_{br}(\omega) = \frac{\pi^2 c^3 \rho(\omega, 1)}{\omega^2}.$$

In the case of equilibrium radiation the brightness temperature coincides with the radiation temperature. In this case T_{br} is proportional to the energy density of radiation.

Let us now consider the non-equilibrium spectrum of radiation. Moreover we assume that the intensity of non-equilibrium radiation in the low-frequency region is considerably higher than that of equilibrium radiation. Then T_{br} will be equal to the temperature of an equilibrium radiation for which the spectral distribution in the low-frequency region coincides with that in a non-equilibrium spectrum. In the case when the intensity in a low-frequency region is much higher than in the equilibrium spectrum, the brightness temperature will be very high, being independent of the mean energy of radiation.

As is clear from (73.14) we have, in ascending order of magnitude

$$|F_{ind}| \sim \sigma u_0 \frac{kT_{br}(\omega)}{mc^2} \beta^2 \sim |F_{sp}| \frac{kT_{br}}{mc^2} \beta^2. \quad (75.1)$$

For equilibrium radiation we always have $|F_{ind}| \ll |F_{sp}|$. On the other hand, if the non-equilibrium spectrum has a brightness temperature T_{br} such that $T_{br} \gg mc^2$, the inequality $|F_{ind}| \gg |F_{sp}|$ is fulfilled.

Now we shall turn our attention to the question of heating the electron gas in the radiation field. The main result arrived at is the existence of a Maxwell distribution with an effective temperature given by formula (74.7).

The physical meaning of this result is that for small momentum transfers in an elementary interaction, the motion of an electron in momentum space is a Brownian motion. After a certain relaxation time, a system of electrons located in an isotropic radiation field reaches a steady state. A random (Gaussian) distribution of momenta is set up for the electrons with an arbitrary distribution of quanta with respect to frequency, $n(\omega)$. In an isotropic field this random distribution is Maxwellian with modulus (temperature) θ . In an equilibrium radiation field, when $n(\omega)$ is given by the Planck formula, from (74.7) we obtain the obvious result $\theta = T$, i.e. the radiation and electron temperatures are equal. If, however, the distribution of quanta is non-equilibrium, the effective electron temperature depends on the form of the function $n(\omega)$.

It is interesting to consider the case where the spectral distribution $n(\omega)$ has a maximum in the region of low frequencies. In this region the occupation numbers $n(\omega)$ are very large compared with unity. Therefore we can let $\hbar \rightarrow 0$

in formula (74.7), i.e. we pass to the classical limit. Thus we obtain

$$\theta \sim \frac{\int n^2(\hbar\omega)^2 \omega^2 d\omega}{4 \int n(\hbar\omega) \omega^2 d\omega} \simeq \frac{\pi^2 c^3}{u_0} \int \frac{(\rho(\omega))^2 d\omega}{\omega^2}. \quad (75.2)$$

In this case, the induced Compton effect dominates the process of electron gas heating. As the process of heating is of a purely classical nature, both formulae (75.2) and (73.8) can be derived in a purely classical manner, i.e. in a random radiation field. In connection with formula (75.2) we observe that:

(1) When the radiation density in the low-frequency region is large enough the electron temperature becomes very high. This means that the average energy of electrons can be much greater than the average energy of photons. It is necessary to emphasize that the heating is due to the action of low-frequency radiation. The energy transfer per unit act of interaction with a photon is very small. But the sheer number of low-frequency photons is so large that the total energy transfer becomes quite considerable.

(2) It is clear that the effective temperature of electrons is of the same order of magnitude as the highest brightness temperature of radiation and always $\theta < T_{br}$.

Therefore we can afford to neglect the effect of frequency change in the process of scattering in the expression for the mobility (74.2). Formula (70.10) for the change of momentum becomes incorrect and formula (74.5) becomes invalid just when the temperature θ rises to the relativistic value $\theta \sim mc^2$. In this case the non-relativistic treatment is inadequate.

(3) When calculating the integral in the numerator of (75.2) it should be taken into account that in the region of very low frequencies the process of bremsstrahlung becomes dominant over that of scattering. Therefore the integral in (75.2) cannot diverge and θ always has a definite value.

The induced scattering has another very surprising and important peculiarity which has never been mentioned before. As is well known the spontaneous scattering cannot occur in homogeneous media or in a spatially homogeneous system. This is easy to see from formula (73.1). If in this formula we let $e \rightarrow 0$ and simultaneously $m \rightarrow 0$ we shall pass to the case of scattering by an infinitely small charge with an infinitesimally small mass. This means that we shall go from the point charge to a continuous homogeneous charged medium. In this limit, we immediately see that the acceleration of an infinitesimal element of media tends to zero. Thus this element will not radiate any secondary waves.

This result is easy to understand from the following simple consideration:

let us consider two parallel infinitely thin layers in homogeneous media. Let us assume that these layers are located at the distance that is equal to half of the wavelength of incident radiation. Then secondary waves radiated by one layer will be fully damped by secondary waves from another layer.

As a matter of fact spontaneous scattering can take place only in inhomogeneous media. In particular this inhomogeneity can arise as a result of fluctuations in homogeneous media.

The situation is quite different in the case of induced scattering. From formula (73.14) for the induced pressure force we see that the acceleration of a unit charge is proportional to $(e/m)^4$. If we pass over to the limit $e \rightarrow 0$ and $m \rightarrow 0$ we obtain the remarkable result: the acceleration of an infinitesimal element of an homogeneous continuous medium in a radiation field is not zero. This means that induced scattering exists even in homogeneous media.

It is possible to interpret this result in the following way: during the act of induced scattering one wave virtually forms inhomogeneity in the medium, the second wave being scattered by this inhomogeneity.

We consider now the range of applicability of the formulae we have obtained. First, however, it is necessary to discuss in more detail the behaviour of an electron in a strong monochromatic wave. By a 'strong' wave we understand the following: in the derivation of the Thomson formula, §36, Part I, it was assumed that the wave was scattered by a mobile electron which acquires a wave velocity $v \sim eE/m\omega \gtrsim c$. The action of the magnetic field on the electron is entirely neglected here. With increasing electric field strength E , the velocity of the electron in the wave increases and can become relativistic, $v \sim c$. The behaviour of the electron in the wave is then radically altered. The magnetic field will give it a rotational motion, so that its complete trajectory will be a closed curve in the form of a figure-of-eight. In this trajectory the electron will radiate electromagnetic waves with various frequencies, differing from the frequency of the incident wave*. The cross section of scattering now turns out to be different from the Thomson cross section, and can be found from simple energy considerations. By (73.1) the energy acquired by an electron per unit volume can be represented in the form $\sigma u_0 c$.

On the other hand, the loss of electron energy in the magnetic field (when $v \sim c$) is given by formula (25.12) of Part II. In the steady state

$$\sigma_0 u_0 c = \left(\frac{d\epsilon}{dt} \right)_{\text{rad}}.$$

* For the radiation spectrum of an electron moving in a magnetic field, see L.D. Landau and E.M. Lifshitz, *Course of theoretical physics*, Vol. 2, *Classical theory of fields*, §74 (Pergamon, Oxford, 1959).

Hence we find for the cross section

$$\sigma \simeq \frac{1}{u_0 c} \left(\frac{d\epsilon}{dt} \right)_{\text{rad}} \simeq \frac{\sigma_T v^2 H^2}{u_0 c^2} \simeq \left(\frac{eE}{m\omega c} \right)^2 \sigma_T = \alpha^2 \sigma_T,$$

where $\alpha \sim eE/m\omega c$. We see that the scattering cross section must increase rapidly with increasing field strength. It should be noted that we have used formula (73.1) for the energy transfer of an electron in its own reference frame. This was possible because the rotational motion of an electron does not cause any increase in energy when averaged over a period.

It is however evident that the cross section cannot rise indefinitely, since the energy acquired by the electron (in unit time) cannot exceed the quantity eEc . This quantity represents the work by the field per sec on an electron moving with the velocity $v = c$. Hence we always have $\alpha^2 \sigma_T c E < eEc$. When $\alpha = (eE/m\omega c) \sim (\lambda/r_0)^{1/2}$, where λ is the wavelength and r_0 the classical electron radius, the cross section begins to decrease as $1/E$. For $\alpha \ll 1$ the Thomson theory of scattering becomes insufficient. The electron in the field of a strong wave cannot be treated as a free particle. These results can be interpreted in the language of quantum mechanics in the following way: the most probable process in a weak wave field is single-quantum scattering, i.e. the ordinary Compton effect.

In the field of a strong wave the most probable process becomes multiple-quantum scattering, in which the number of quanta in the act of scattering is not conserved (frequencies different from that of the scattered wave are radiated).

Now we shall pass to a discussion of the application of the theory developed above. For the case of the spectral radiation field the first, obvious, condition for the applicability of the theory is the requirement that the characteristic time of energy accumulation by an electron should be small compared with the inverse of the characteristic frequency of the spectrum. Otherwise the radiation field cannot be a random field for the electron.

We estimate this characteristic time, which is equivalent to the retardation time for an electron in a spectral radiation field.

Assuming for simplicity that the latter is isotropic, we have, if $v \sim c$

$$\tau |F_{\text{ind}}| \lesssim mc. \quad (75.3)$$

Hence we have the inequality

$$\frac{mc}{|F_{\text{ind}}|} > \frac{1}{\omega}. \quad (75.4)$$

Or, using the formulae for $|F_{\text{ind}}|$ we obtain the equivalent inequality $\alpha < 1$. The last inequality is identical with the condition for the applicability of the ordinary theory of spontaneous scattering.

The second condition for the applicability of the theory of induced scattering is the inequality

$$|F_{\text{ind}}|c < eEc. \quad (75.5)$$

Substituting the value of F_{ind} we again arrive at inequality (73.4). This result shows that in the case of induced scattering there cannot be an increase in the cross section compared with the Thomson cross section, irrespective of the value of the parameter $\alpha = eE/mc\omega$. It must be emphasised that the above two conditions for the applicability of the theory of induced scattering in a spectral field are completely independent. Whereas the first of them expresses the essential character of the spectral field, the second relates to the existence of a limiting velocity of propagation of interactions ($v \leq c$). When the parameter $(eE/m\omega c)$ tends to unity, the cross section of induced radiation begins to decrease with the field according to the law

$$\sigma_{\text{eff}}^{\text{ind}} \sim \frac{1}{E^2}.$$

Thus the induced reaction force attains its maximum possible value when $\alpha = (eE/mc\omega) \sim 1$.

There is a series of applications of the preceding theory in astrophysics, which we are unable to discuss within the framework of this book*.

The examples we have given will enable the reader to see how, in the presence of high-intensity fields, even in the simplest system, quite new specific phenomena appear — phenomena which are wholly different from those that arise in weak fields.

At the same time the reader will be convinced of the inexhaustibility of classical physics and the great power of the methods of quantum theory.

* See E. Levich and R. Sunaev, *Astroph. Lett.* 7 (1971) 69.



SUBJECT INDEX

- Absorption energy, 225
- Acceptor, 357
- Activity, 11
- Anisotropic metal, 333
- Antiferromagnet, 286

- Band, energy, 250
- Band width, 251
- Bardeen-Cooper-Schrieffer theory, 341
- Barodiffusion coefficient, 39
- Barrier, thermal transport through, 56
- Basic kinetic equation, derivation of, 58
 - - - in closed system, 64
 - - - in quasi-closed system, 65
 - - -, validity of, 61
- Basis vector, 233
- Bloch wave, 244
- Bogoliubov approximation, 26
 - transformation, 28
- Boltzmann's equation, 77
 - - and Liouville's equation, 85
 - -, Chapman-Enskog solution of, 109, 115
 - - for a plasma, 166
 - - for a plasma with collisions, 179
 - - for electrons in a solid, 300, 327
 - - for electrons in a solid, solution of, 329
 - - for neutrons in matter, 153
 - - for polyatomic gases, 145
 - - for quasi-gaseous system, 139
 - -, linear, 139
 - -, linear, in external field, 142
 - -, linear, solution of, 142
 - -, linearized, 110
 - -, method of moments solution of, 106
 - -, relaxation approximation of, 124
- Bose-Einstein condensation, 21, 23
 - - - in phonon gas, 379
 - - - temperature, 21
- Bose gas, degenerate ideal, 18
 - -, Hamilton operator of, 25
 - -, imperfect, 24
- Bosons, partition function of ideal gas of, 16
- Bragg reflection, 256
- Bravais lattice, 233
- Bremsstrahlung, free-free, 370
- Brightness temperature, 388
- Brillouin zone, 244
- Brownian motion, 56
- Bulk viscosity coefficient, 43

- Chapman-Enskog method, 109, 115
- Charge density in plasma, 263
- Chemical reaction rate, 58
- Classical approximation, quasi-, 14
- Closed system and kinetic equation, 64
- Cohesive energy, 233
 - force, 230
- Collision integral, 74, 76, 78
 - - for electrons in a photon gas, 372
 - - for electrons in a solid, 301, 326, 332
 - invariant, 92
 - time, 83
- Compton scattering, 368
- Condensation and superfluidity, Bose, 32
 - temperature, Bose-Einstein, 21
- Conduction band, 270
 - , electron, 299
 - , equation of thermal, 46, 49
- Conductivity, Boltzmann's versus correlation function expression for, 206
 - , electrical, 205, 302, 304, 324

- , -, in light wave field, 311
- , -, in magnetic field, 310
- , -, in semiconductor, 362
- , -, tensor in plasma, 172
- , thermal, 46
- Conductor, ideal, 305
- Conservation law, energy, 44, 97
 - , mass, 37, 40, 93
 - , momentum, 42, 95
- Continuity equation, 38
- Continuous medium approximation, 37
- Convective diffusion, equation of, 40
- Correlation function, 79, 195, 203, 227
- Cooper interaction, 335
 - pair, 339
- Cross-section, differential scattering, 74
 - , transport scattering, 130
- Crystal, covalent, 233
 - , ionic, 233
 - lattice, 232
 - - geometry, 233
- Curie point, 285
- Current, electric, 205
 - , -, in plasma, 171
- Damping of plasma waves, 177, 178
- Debye approximation, 167
- Deformation rate tensor, 116
- De Haas-Van Alphen effect, 280
- Density distribution for particles in matter, 166
 - matrix, 5
 - , time evolution of, 7
- Detailed balance, principle of, 61
- Diamagnetic susceptibility, 280
- Diamagnetism, orbital, 275
- Dielectric constant of quantum plasma, 261, 264
 - crystal, 351
 - permeability tensor, 173
 - susceptibility, 226
- Diffusion coefficient, 131, 203, 221
 - , baro-, 39
 - , generalized, 54
 - , molecular, 39
 - - of electrons in radiation field, 387
 - , thermal, 133, 135
 - , -, coefficient, 134
 - equation, generalized, 56
 - flux, 38, 94
 - of light particles through a gas of heavy particles, 126
- Dispersion law, 238
 - of longitudinal plasma waves, 174, 177
 - of transverse plasma waves, 178
 - relation for plasma waves, 267
- , sound, 136
- Dissipative function, 214
- Distribution function, 73, 80
 - , basic kinetic equation for, 82
 - , Chapman-Enskog forms for, 119, 122
 - , electron equilibrium, 328
 - , generalized, 50
 - - in perturbed system, 192
 - , phonon equilibrium, 328
- Donor, 357
- Dynamic perturbation, 191
- Effective mass, 249, 269
 - wave vector, 243
- Electrical conductivity, 204, 206, 302, 304, 324
 - - in magnetic field, 310
 - current, 205
- Electron as a quasi-particle, 269
 - conduction, 299
 - distribution function in radiation field, 386
 - equilibrium distribution function, 328
 - free path length, 304, 318
 - - - due to phonon collision, 318, 321, 324, 333
 - gas, free energy of, 272
 - - and free phonon gas, 368
 - - susceptibility, 280
 - group velocity, 257
 - in crystal, energy spectrum of:
 - in tight binding approximation, 249
 - in nearly free electron approximation, 255
 - - , Schrödinger equation for, 242
 - - , wave function of, 248
 - in radiation field, 366, 380
 - - - , diffusion coefficient for, 387
 - - - , force on, 383, 386

- interaction, screened, 268
- pair, 335, 339
- phonon interaction, 269, 287
- - - Hamiltonian, 288
- Electrons in photon gas, kinetic equation for, 373
- , number of free, 269
- Electron thermal conductivity, 325
- velocity in electric field, 258
- Energy band, 250
- conservation law, 44, 97
- flux density, 45, 98
- gap in superconductor, 341, 344
- Entropy and negative temperature, 69
- balance equation, 47
- , definition of, 99
- flow density, 48
- , increase of, 99
- , mean, 200
- production, 48
- - in stationary state, 219
- Enskog's equation, 91
- Equation of state, 11
- - - for ideal Bose gas, 17
- - - for ideal Fermi gas, 17
- Equilibrium distribution, 101
- , local, 103
- , establishment of, 105
- , - - in ion-electron plasma, 185
- , local, 36
- , local statistical, 200
- Euler's equation, 114
- Exchange integral, sign of, 286
- interaction, 233, 283
- Exciton, 241, 352
- energy, large radius, 353
- -, small radius, 352
- Extrinsic semiconductor, 356
- -, distribution function of, 359
- -, n-type, 357
- -, p-type, 357
- Fermi distribution, 301
- fluid, 344
- -, kinetic properties of, 348
- -, quasi-particles in, 346
- -, thermal conductivity of, 348
- -, thermodynamic quantities of, 347
- -, viscosity of, 348
- Fermions, partition function of ideal gas of, 16
- Ferromagnetism, 281
- , model for, 281
- Feynman diagram, 296
- Fick's law, 40
- Fluctuation-dissipation theorem, 225
- Flux density in quasi-gaseous system, 142
- , generalized, 197
- Fokker-Planck equation, 54, 372
- - -, second form of, 55
- Fowler's law, 317
- Free path, mean transport, 324
- - length, electron, 304, 318
- - -, -, due to phonon collisions, 321, 324, 333
- Free electron gas and free phonon gas, 368
- - - in radiation field, 365
- Free-free bremsstrahlung, 370
- Frenkel exciton, 352
- Fröhlich Hamiltonian, 294
- Generalized transport equation, 91
- Gibbs distribution, 66
- -, local equilibrium, 200
- Group velocity, electron, 257
- Hall constant, 310
- - in semiconductor, 363
- effect, 305
- - in semiconductor, 363
- Heat flux density, 46
- Hermite-Sonine polynomials, 107
- Hole, 345
- , energy of, 346
- Hydrodynamic time, 83
- Ideal Bose gas, degenerate, 18
- - -, isotherms of, 23
- gas, 11, 72
- Impurity, 299
- Incompressible medium, 38, 48
- Insulator, 270
- Interaction, effective quantum, 14
- Intrinsic semiconductor, 356
- -, distribution function of, 357

- Ions, size effect of, 258
- Irreversibility, 62, 89, 195
- Irreversible process, 33
 - —, quasi-macroscopic treatment of, 34
 - —, kinetic approach of, 34
- Isotope effect, 335

- Kinetic coefficient, 37
 - equation, basic, derivation of, 58
 - —, —, for closed system, 64
 - —, —, for quasi-closed system, 65
 - —, —, for the distribution function, 83
 - —, —, validity of, 61
 - — and rotational motion, 146
 - — for electrons in a solid, 300, 327
 - — for electrons in a solid, solution of, 329
 - — for electrons in photon gas, 371
 - — for photons, 373
 - — for polyatomic gases, 145
 - — with self-consistent field, 169

- Landau damping, 177
- Landau's kinetic equation, 182, 184
 - — — and the H -theorem, 185
 - — — and the Maxwell distribution, 184
- Langmuir frequency, 177
- Laser, 71
- Lattice, Bravais, 233
 - constants, 234
 - , crystal, 232
 - , reciprocal, 234
 - vibrations, 235
 - with a basis, 234
- Level population, 68
- Liouville's equation, 80
 - — and Boltzmann's equation, 85
- Local equilibrium, 36
 - — distribution, 103
 - — — for polyatomic gases, 147
 - Gibbs distribution, 200
 - statistical equilibrium, 200
- Lorentz method, 127

- Magnetic flux, quantization of, 341
 - susceptibility, 274
- Magnetization, spontaneous, 285
- Magnon, 241, 283
 - distribution function, 284
- Magnons, number of, 285
- Markov chain, 51
- Maser, 71
- Mass concentration, 39
 - conservation law, 37, 40, 93
- Matter, motion of particles in, 151
 - , neutron density distribution in, 165
- Maxwellian molecules, 108
- Maxwell-Lorentz equations in a plasma, 168
- Maxwell's distribution, 102
 - —, local, 103
- Mean, quantum mechanical, 5
 - , statistical, 5, 8, 10
- Metal, 270
 - , anisotropic, 333
 - , good, 269
 - , magnetic properties of, 271
 - , optical properties of, 310
- Metalloid, 269
- Microscopic relaxation time, 83
- Mobility vector, 54, 373
- Moderation of neutrons, 151
- Molecular diffusion coefficient, 39
- Moment approximation, thirteen, 108
- Moments solution of Boltzmann's equation, method of, 106
- Momentum conservation law, 42, 95

- Navier-Stokes equation, 44, 120
- Nearly free electron approximation, 251, 256
- Negative temperature, 69
 - — and entropy, 69
 - — and radiation, 71
 - —, example of system with, 69
- Neutrons in gas of heavy nuclei, energy distribution of, 157
 - in hydrogen, energy distribution of, 155
 - in matter, Boltzmann's equation for, 153
 - — —, density distribution of, 165
 - — —, energy distribution of, 153
 - — —, spatial distribution of, 161
 - — — with isotropic scattering, 162
- Newtonian fluid, 42

- Newton's law of viscosity, 43
 Nyquist's formula, 228

 Onsager's hypothesis, 209
 — relations, 213
 — — and microscopic reversibility, 214
 — theory, 208
 Optical properties of a metal, 310
 Orbital susceptibility, 280

 Paramagnetic gas, thermal conductivity of, 150
 Paramagnetism, spin, 274
 Particles in matter, density distribution of, 166
 — — —, motion of, 151
 Partition function, canonical, 9
 — —, grand, 10
 — — of ideal gas, 13
 Pauli equation, 275
 Peltier heat, 224
 Perturbation, dynamic, 191
 —, periodic, 194
 —, thermal, 191, 199
 Phonon, 238
 — absorption probability, 296
 — annihilation, 290
 — — operator, 239
 — creation, 290
 — — operator, 239
 — — probability, 295
 — electron interaction, 287
 — — — Hamiltonian, 290
 — equilibrium distribution function, 328
 — momentum, 239
 —, virtual, 297
 Photoelectric current, 316
 — effect, 312
 — —, surface, 313
 — —, volume, 313
 Photon distribution in electron gas, 376
 — gas, Bose-Einstein condensation in, 379
 — —, cooling of, 377
 — —, heating of, 377, 380
 — —, kinetic equation for, 373
 — spectral distribution, 370
 Planck formula, 67

 Plasma, Boltzmann's equation for, 166
 —, degenerate quantum, 260
 —, dielectric permeability tensor in, 173
 —, electric conductivity tensor in, 172
 —, electric current in, 171
 —, electron-ion, 185
 —, — —, relaxation times in, 188
 —, Maxwell-Lorentz equations for, 168
 —, self-consistent field, approximation for, 168
 —, — —, solution for, 169
 — wave, 267
 — —, damping of, 169, 177, 178
 — —, dispersion of, 169, 174, 178
 — —, longitudinal, 177
 — —, transverse, 178
 — with collisions, Boltzmann's equation for, 179
 Plasmon, 241, 267
 Point charge in plasma, field of, 265
 Poisson bracket, 80
 Polarization vector, 238
 Polaron, 241
 Polyatomic gas, kinetic equation for, 145
 — —, thermal conductivity of, 147
 Pressure of ideal Bose gas, 22
 — thermomolecular, 217
 Probabilities, hypothesis of equal, 9
 Probability distribution, 9
 Pseudopotential, 259

 Quantization of magnetic flux, 341
 Quasi-closed system, 4
 — — — and basic kinetic equation, 65
 Quasi-equilibrium, 35
 Quasi-gaseous system, 138
 — — —, flux density in, 142
 — — —, in external field, 142
 Quasi-momentum, 240
 Quasi-particle, 29, 239, 269
 — — in Fermi fluid, 347
 — — lifetime, 241

 Radiating particle, force on, 380, 383
 Radiation reaction force, in anisotropic field, 385
 — — —, induced, 384
 Random phase condition, 8, 61

- Reciprocal lattice, 234
- Relaxation, 34
 - approximation of Boltzmann's equation, 123
 - time, microscopic, 83
- Resistance, residual, 305, 325
- Response function, 195
 - of a system, classical, 193
 - - - - , quantum mechanical, 199
 - to a thermal perturbation, 202
- Rotational motion and kinetic equation, 146
- Scattering cross-section, differential, 74
 - - - , transport, 130
- Screening, 265
 - and electron interaction, 268
 - length, 265
- Seebeck effect, 224
- Semiconductor, 271, 355
 - , conductivity of, 362
 - , extrinsic, 356
 - , - , distribution function, 359
 - , - , n-type, 357
 - , - , p-type, 357
 - , Hall constant of, 363
 - , Hall effect in, 363
 - , intrinsic, 356
 - , - , distribution function, 357
- Senftleben effect, 149
- Shear viscosity coefficient, 43
- Smoluchowski equation, 51
- Solid state, adiabatic approximation in, 231
 - - Hamiltonian, 231, 294
 - - model, 230
 - - wave function, 231
- Sound dispersion, 136
 - velocity, 137
- Spectral distribution, phonon, 370
- Spin paramagnetism, 274
- Spin wave, 283
- Stationary non-equilibrium process, 218
 - state, 35
- Statistical matrix, 5
 - operator, 5
 - - , canonical, 10
 - - , grand canonical, 10
 - - in perturbed system, 198
 - - , time evolution of, 7
- Stress, 41
 - tensor, 41, 96
- Subsystem, closed, 4
 - , open, 4
 - , quasi-closed, 4
- Superconductivity, 334, 343
- Superconductor in magnetic field, 339
 - , model Hamiltonian for, 341
- Superfluidity, 30
 - and Bose condensation, 32
 - , condition for, 31
- Susceptibility, diamagnetic orbital, 280
 - , generalized complex, 194, 226
 - of electron gas, 280
 - , paramagnetic, 274
- Temperature, negative, 69
- Thermal conduction, equation of, 46, 49
 - conductivity, 46, 122
 - - , electronic, 325
 - - of diatomic gas, 147
 - - of Fermi fluid, 348
 - - of paramagnetic gas, 149
 - diffusion, 133, 135, 220
 - - coefficient, 134, 222
 - diffusivity, 49
 - perturbation, 191, 199
 - wavelength, 13
- Thermodynamic force, 201, 209, 216
- Thermoelectric effect, 222
- Thermoelectromotive force, 224
- Thermomechanical effect, 218
- Thermomolecular pressure, 217
- Thirteen-moment approximation, 108
- Thomson cross-section, 374, 393
 - formula, 370, 392
 - relation, second, 224
- Tight binding approximation, 245, 251
- Time correlation function, 195, 203, 227
- Transition metals, 281
- Transport coefficient, generalized, 196, 203
 - - , Onsager expression, 211
 - - , tensor of, 197
 - equation, 49
 - - , generalized, 91

- mean free path, 132, 324
- scattering cross-section, 130
- through a barrier, thermal, 56
- Umklapp process, 289
- Unit cell, 233
- Valence electron, 230
- Velocity of electron in electric field, 258
- Virtual phonon, 297
- Viscosity coefficient, bulk, 43
 - -, first, 43
 - - of an ideal gas, 119
- -, relaxation time approximation for, 125
- -, second, 43
- -, shear, 43
- of a Fermi fluid, 348
- Wannier exciton, 353
- Wave vector, effective, 243
- Wiedemann-Franz law, 325
- Zero point oscillations in a plasma, 268
- Zero sound, 348
 - - velocity, 350